

PENDEKATAN *CART* UNTUK MENDAPATKAN FAKTOR YANG MEMPENGARUHI TERJANGKITNYA PENYAKIT DEMAM TIFOID DI ACEH UTARA

Muhammad Sjahid Akbar¹, Dina Yuanita², dan Sri Harini³

^{1,2}Jurusan Statistika ITS

³Jurusan Matematika, UIN Malulana Malik Ibrahim Malang

e-mail: sri_harini21@yahoo.com

ABSTRACT

Typhoid fever is a disease caused by Salmonella typhi bacteria. It is attack the digestive tract. Typhoid fever caused by poor sanitation and personal hygiene is not good. According to the Basic Health Research in 2007 showed that the prevalence of typhoid fever in Indonesia of 1.6%. NAD Province is hight typhoid fever prevalence(2,96 %). Because Having traced the biggest contributor was derived from NAD. Therefore, the research conducted to find factors that influence the outbreak of typhoid fever in NAD. research using the CART Method. The results of the analysis indicate that the main factor causing typhoid fever was drinking water reservoirs. The other factors are waste water reservoirs, the physical quality of drinking water, a habit washing hands with soap before eating, the bowel, the dump, gender, socioeconomic status, habits of washing hands with soap after defecation and health education.

Keywords: *CART, Typhoid fever*

PENDAHULUAN

Penyakit demam tifoid seringkali menjadi sebab seseorang harus menjalani rawat inap. Demam Tifoid atau *Typhoid fever* yang biasa juga disebut typhus atau types oleh orang awam, merupakan penyakit yang disebabkan oleh bakteri *Salmonella Typhi* (*S. Typhi*). Bakteri *S. Typhi* menyerang bagian saluran pencernaan. Puslitbang Sistem dan Kebijakan Kesehatan menyatakan demam tifoid disebabkan pencemaran air minum dan sanitasi yang buruk. Demam tifoid adalah penyakit infeksi akut yang menyerang mulai dari usia balita, anak-anak dan dewasa. Data *World Health Organization (WHO)* tahun 2003 memperkirakan terdapat sekitar 17 juta kasus demam tifoid di seluruh dunia dengan kejadian 600.000 kasus kematian tiap tahun (Anonim, 2008). Angka kejadian demam tifoid diketahui lebih tinggi pada negara berkembang khususnya di daerah tropis. Sehingga tak heran jika demam tifoid banyak ditemukan di Indonesia. Hasil Riset Dasar Kesehatan tahun 2007 menunjukkan bahwa persentase penduduk yang terjangkit demam tifoid dibandingkan dengan seluruh penduduk (prevalensi) di Indonesia sebesar 1,6%. Provinsi NAD merupakan prevalensi tifoid tertinggi yaitu sebesar 2,96%. Setelah ditelusuri ternyata penyumbang terbesar berasal dari kabupaten Aceh Utara. Oleh karena itu penelitian dilakukan di wilayah Aceh Utara untuk mendapatkan faktor-faktor yang menyebabkan terjangkitnya penyakit demam tifoid.

Ada tiga penelitian yang digunakan untuk dasar penelitian ini. Tugas akhir Nunik Hidayati mahasiswa S1 jurusan Statistika FMIPA ITS, thesis Rahayu Lubis mahasiswa pasca sarjana jurusan Kesehatan Masyarakat di Universitas Sumatera Utara, dan penelitian Bambang Wasito Tjipto peneliti dari Puslitbang System dan Kebijakan Kesehatan. Hidayati (2001) memodelkan kasus penyakit demam tifoid di Jawa Timur dengan menggunakan regresi poisson. Asumsi yang harus dipenuhi apabila menggunakan metode regresi poisson adalah variabel dependen harus diskrit dan berdistribusi poisson. Ada beberapa faktor resiko yang diduga mempengaruhi terjangkitnya penyakit demam tifoid antara lain kepadatan penduduk, prosentase cakupan penduduk pemakai air bersih, prosentase cakupan penduduk pemakai jamban keluarga, prosentase kondisi rumah yang memenuhi syarat, prosentase cakupan pembuangan sampah sementara yang memenuhi syarat, prosentase cakupan tempat pengolahan makanan yang memenuhi syarat, dan prosentase cakupan penduduk pemakai sarana pembuangan air limbah. Hasil dari penelitian Hidayati variabel yang mempengaruhi terjangkitnya demam tifoid adalah kepadatan penduduk, prosentase cakupan penduduk pemakai air bersih, prosentase cakupan pembuangan sampah sementara yang memenuhi syarat, prosentase cakupan tempat pengolahan makanan yang memenuhi syarat, dan prosentase cakupan penduduk pemakai sarana pembuangan air limbah. Sedangkan penelitian Lubis (2007) mempelajari faktor risiko yang

berhubungan dengan kejadian penyakit demam tifoid pada penderita yang dirawat di RSUD Dr. Soetomo Surabaya dengan menggunakan regresi logistik. Variabel yang digunakan seperti tingkat pengetahuan, higiene perorangan, kebiasaan makan/minum diluar rumah dan sanitasi lingkungan. Hasilnya faktor yang mempengaruhi kejadian penyakit demam tifoid adalah *hygiene* perorangan dan kualitas air minum. Selain itu, Tjipto (2009) meneliti faktor-faktor yang berpengaruh terhadap kejadian penyakit demam tifoid pada balita di Indonesia dengan analisis multivariate logistik biner. Tjipto (2009) menyatakan bahwa demam tifoid erat kaitannya dengan higiene perorangan dan sanitasi lingkungan. Hasil penelitian menunjukkan bahwa faktor-faktor yang berpengaruh agar tidak terjadi penyakit infeksi tifoid adalah buang air besar ditempat yang baik (jamban), dan mencuci tangan dengan benar (memakai sabun).

Tujuan penelitian ini adalah untuk mendapatkan faktor yang mempengaruhi terjangkitnya demam tifoid menggunakan metode *Classification and Regression Trees* (CART). Alasan menggunakan metode CART adalah CART merupakan salah satu metode non parametrik dengan hasil analisis berupa topologi pohon atau berupa grafis sehingga hasil analisis lebih mudah diinterpretasi (Lewis dan Roger, 2000). Data yang digunakan merupakan data sekunder yang diambil dari Riskesdas tahun 2007 dan Susenas tahun 2007. Data dihimpun oleh Badan Litbangkes Departemen Kesehatan RI. Total sampel ART di Aceh Utara adalah sebanyak 2.491 ART. Pada penelitian ini data yang digunakan 1816 data ART dengan batasan ART minimal berusia 10 tahun. Variabel respon yang digunakan berskala biner yaitu, 1 untuk anggota rumah tangga terinfeksi demam tifoid dan 2 untuk anggota rumah tangga yang tidak terinfeksi demam tifoid. Sedangkan variabel prediktor yang digunakan dalam penelitian ini adalah. asal daerah (x_1), jenis kelamin (x_2), status sosial ekonomi (x_3), kualitas fisik air minum (x_4), tempat penampungan air minum (x_5), tempat pembuangan sampah (x_6), tempat penampungan air limbah (x_7), tempat buang air besar (x_8), kebiasaan cuci tangan pakai sabun setelah buang air besar (x_9), kebiasaan cuci tangan pakai sabun sebelum makan (x_{10}), dan penyuluhan kesehatan (x_{11}).

CLASSIFICATION AND REGRESSION TREES (CART)

Classification and Regression Trees (CART) adalah suatu metode teknik pohon keputusan (Breiman et al., 1993). *CART* menghasilkan suatu

pohon klasifikasi jika variabel responnya kategorik, dan menghasilkan pohon regresi jika variabel responnya kontinu. Tujuan utama *CART* adalah untuk mendapatkan suatu kelompok data yang akurat sebagai penciri dari suatu pengklasifikasian. Klasifikasi pohon dalam *CART* melibatkan 4 komponen, yaitu variabel respon, variabel prediktor, data *learning*, dan data *testing*. Data *learning* untuk verifikasi model dan data *testing* untuk validasi model.

Sebagai ilustrasi struktur pohon klasifikasi dapat dilihat pada Gambar 1. Simpul utama dinotasikan dengan t_1 sedangkan *internal nodes* (simpul dalam) dinotasikan dengan t_2, t_3, t_4, t_7, t_9 dan t_{13} . Simpul akhir atau simpul terminal adalah $t_5, t_6, t_8, t_{10}, t_{11}, t_{12}, t_{14}$ dan t_{15} . Penghitungan *depth* (kedalaman) pohon dimulai dari simpul utama t_1 yang berada pada kedalaman 1, sedangkan t_2 dan t_3 berada pada kedalaman 2 begitu seterusnya sampai pada t_{14} dan t_{15} yang berada pada kedalaman 6.

Pembentukan pohon klasifikasi terdiri atas 3 tahap yang memerlukan *learning sample L*. Tahap pertama adalah pemilihan pemilah. Menurut Breiman et al. (1993) setiap pemilahan hanya bergantung pada nilai yang berasal dari satu variabel independen. Rumus kemungkinan pemilah disajikan sebagai berikut.

$$\begin{aligned} \text{Variabel independen kontinu} &= n - 1 \text{ pemilahan} \\ \text{Variabel independen kategori nominal} &= 2^{L-1} - 1 \text{ pemilahan} \\ \text{Variabel independen kategori ordinal} &= L - 1 \text{ pemilahan} \end{aligned} \tag{1}$$

Setelah semua kemungkinan pemilah didapatkan, masing-masing pemilah dicari nilai *goodness of split*. *Goodness of split* merupakan suatu evaluasi pemilahan oleh pemilah s pada simpul t . *Goodness of split* $\emptyset(s, t)$ didefinisikan sebagai penurunan keheterogenan. Sehingga semakin besar nilai *goodness of split* semakin homogen simpul anak yang dihasilkan.

$$\emptyset(s, t) = \Delta i(s, t) = i(t) - P_L i(t_L) - P_R i(t_R) \tag{2}$$

Pengembangan pohon dilakukan dengan mencari semua kemungkinan pemilah pada simpul t_1 sehingga ditemukan pemilah s^* yang memberikan nilai penurunan keheterogenan tertinggi yaitu,

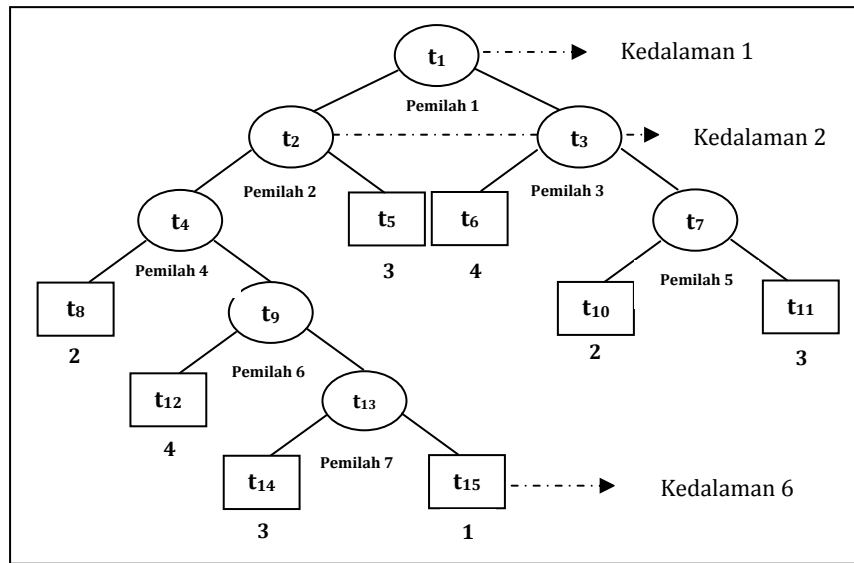
$$\Delta i(s^*, t_1) = \max_{s \in S} i(s, t_1) \tag{3}$$

Dengan $i(t)$ adalah fungsi keheterogenan indeks Gini, $\emptyset(s, t)$ adalah kriteria *goodness of split*, $P_L i(t_L)$ adalah proporsi pengamatan dari simpul t menuju simpul kiri, dan $P_R i(t_R)$ adalah proporsi pengamatan dari simpul t menuju simpul kanan.

Indeks Gini sebagai metode pemilahan yang digunakan mempunyai fungsi sebagai berikut.

$$i(t) = \sum_{i=j} p(i|t)p(j|t) \quad (4)$$

dengan, $p(i|t)$ adalah proporsi kelas i pada simpul t , dan $p(j|t)$ adalah proporsi kelas j pada simpul t .



Gambar 1 Struktur Klasifikasi Pohon

Tahap kedua adalah penentuan simpul terminal (penghentian pembentukan pohon). Simpul t dapat dijadikan simpul terminal jika (1) Tidak terdapat penurunan keheterogenan yang berarti. (2) Hanya terdapat satu pengamatan ($n=1$) pada tiap simpul anak. (3) Adanya batasan minimum n /pengamatan pada simpul anak. dan (4) Adanya batasan jumlah level atau tingkat kedalaman pohon maksimal (Lewis, 2000).

Tahap ketiga adalah penandaan label tiap simpul terminal berdasarkan aturan jumlah anggota kelas terbanyak, yaitu:

$$p(j_0|t) = \max_j p(j|t) = \max_j \frac{N_j(t)}{N(t)} \quad (5)$$

dengan $p(j|t)$ adalah proporsi kelas j pada simpul t , $N_j(t)$ adalah jumlah pengamatan kelas j pada simpul t , dan $N(t)$ adalah jumlah pengamatan pada simpul t . Label kelas simpul terminal t adalah j_0 yang memberi nilai dugaan kesalahan pengklasifikasian simpul t terbesar.

Setelah terbentuk pohon maksimal tahap selanjutnya adalah pemangkasan pohon untuk mencegah terbentuknya pohon klasifikasi yang berukuran sangat besar dan kompleks. Sehingga diperoleh ukuran pohon yang layak berdasarkan *cost complexity pruning*. Besarnya *resubstitution estimate* pohon T pada parameter kompleksitas α yaitu :

$$R\alpha(T) = R(T) + \alpha|\bar{T}| \quad (6)$$

dengan $R\alpha(T)$ adalah *resubstitution* suatu pohon T pada kompleksitas α , $R(T)$ adalah *resubstitution estimate*, α adalah *Parameter cost - complexity* bagi penambahan satu simpul akhir pada pohon

T , dan $|\bar{T}|$ adalah banyaknya simpul terminal pohon T . *Cost complexity pruning* menentukan pohon bagian $T(\alpha)$ yang meminimumkan $R\alpha(T)$ pada seluruh pohon bagian untuk setiap nilai α . Nilai parameter kompleksitas α akan secara perlahan meningkat selama proses pemangkasan. Selanjutnya pencarian pohon bagian $T(\alpha) < T_{max}$ yang dapat meminimumkan $R\alpha(T)$ yaitu :

$$R\alpha(T(\alpha)) = \min_{T < T_{max}} R\alpha(T) \quad (7)$$

Setelah dilakukan pemangkasan diperoleh pohon klasifikasi optimal yang berukuran sederhana namun memberikan nilai pengganti yang cukup kecil. Penduga pengganti yang sering digunakan adalah penduga sampel uji (*test sample estimate*) dan validasi silang lipat V (*Cross Validation V-Fold Estimate*). Menurut Breiman *et al.* (1993) jika jumlah sampel yang digunakan lebih kecil dari 3000 pengamatan penduga pengganti yang digunakan adalah *Cross Validation V-Fold Estimate*.

Penduga validasi silang lipat V sering digunakan apabila amatan yang ada tidak cukup besar. Amatan dalam L dibagi secara acak menjadi V bagian yang saling lepas dengan ukuran kurang lebih sama besar untuk setiap kelasnya. Pohon $T^{(v)}$ dibentuk dari $L-L_v$ dengan $v = 1, 2, \dots, V$. Misalkan $d^{(v)}(x)$ adalah hasil pengklasifikasian. Penduga sampel uji untuk $R(T_t^{(v)})$ yaitu

$$R^{ts}(T_t^{(v)}) = \frac{1}{Nv} \sum_{(x_n, j_n) \in L_v} X(d^{(v)}(x_n) \neq j_n) \quad (8)$$

dengan $Nv = N/V$ adalah jumlah amatan dalam L_v . Kemudian dilakukan prosedur yang sama

menggunakan seluruh L , maka penduga validasi silang lipat V untuk $T_t^{(V)}$ adalah :

$$R^{Cv}(T_t) = \frac{1}{V} \sum_{v=1}^V R^{ts}(T_t^{(v)}) \quad (9)$$

Pohon klasifikasi optimum dipilih T^* dengan

$$R^{Cv}(T^*) = \min_t R^{Cv}(T_t) \quad (10)$$

APLIKASI DAN PEMBAHASAN

Penelitian menggunakan variabel respon kategorik berskala biner. Bernilai 1 untuk anggota rumah tangga terinfeksi demam tifoid dan 0 untuk anggota rumah tangga yang tidak terinfeksi demam tifoid, sehingga didapatkan pohon klasifikasi untuk menjelaskan keterkaitan 11 variabel prediktor yang diduga mempengaruhi terjangkitnya penyakit demam tifoid. Pada klasifikasi pohon data sampel anggota rumah tangga terjangkau dan tidak terjangkau demam tifoid di Aceh Utara dibagi menjadi dua kelompok yaitu data *learning* dan data *testing*. Penelitian ini menggunakan perbandingan data *learning* 75% (1.362 data) dan *testing* 25% (454 data).

Tahap pertama pembentukan pohon klasifikasi maksimal adalah pemilah-pemilah. Perhitungan pemilah pada setiap variabel prediktor menggunakan persamaan (1). Hasil yang diperoleh adalah variabel asal daerah, variabel jenis kelamin, variabel status sosial ekonomi, variabel kualitas air minum, variabel tempat buang air besar, variabel kebiasaan cuci tangan pakai sabun setelah buang air, variabel kebiasaan cuci tangan pakai sabun sebelum makan, dan variabel keikutsertaan penyuluhan dengan 1 kemungkinan pemilahan. Variabel kondisi penampungan air minum dan variabel kondisi tempat pembuangan sampah dengan 3 kemungkinan pemilahan. Dan variabel kondisi penampungan air limbah dengan 15 kemungkinan pemilahan. Penelitian ini menggunakan metode pemilahan Indeks Gini sesuai persamaan (4). Pemilah terbaik adalah pemilah yang menghasilkan nilai penurunan keheterogenan tertinggi (kriteria pemilahan *goodness of split* pada persamaan (3)). Pemilah terbaik pada simpul 1 (pemilah utama) pada penelitian ini adalah variabel tempat penampungan air (x_5). Variabel tempat penampungan air terpilih sebagai pemilah utama karena menghasilkan nilai penurunan keheterogenan tertinggi pada simpul 1 (Gambar 2). Informasi hasil perhitungan penurunan keheterogenan pada setiap pemilah di simpul 1 disajikan pada Table 1.

Tahap kedua yaitu penentuan simpul terminal. Simpul t dikatakan sebagai simpul terminal jika tidak terdapat penurunan keheterogenan yang berarti sehingga tidak akan dipilah lagi. Simpul terminal adalah simpul yang berwarna merah, biru dan putih. Pohon klasifikasi maksimal (*maximal tree*) dari data anggota rumah tangga yang terjangkau maupun tidak terjangkau demam tifoid ditunjukkan pada Gambar 2.

Tabel 1. Nilai Penurunan Kehlerogenan Variabel Pemilah pada Simpul 1

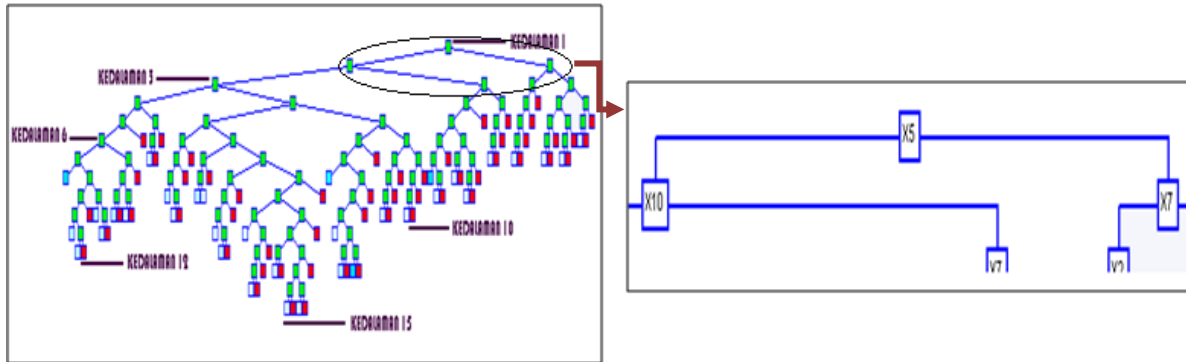
Pemilah	Split	$\Delta i(s, t)$
X ₁	1	0,000418599
X ₂	1	0,000126786
X ₃	1	0,000714858
X ₄	1	0,000867109
X ₅	1,2	0,001200313
X ₅	1,3	2,08527E-05
X ₅	2,3	0,000896645
X ₆	1,2	9,47417E-05
.	.	.
X ₇	2,3	0,000400087
.	.	.
.	.	.
X ₈	1	0,000343279
X ₉	1	0,000118855
X ₁₀	1	0,001019946
X ₁₁	1	0,00041325

Pohon klasifikasi maksimal terdiri dari 89 simpul terminal dengan 15 kedalaman. Kedalaman adalah jumlah level atau tingkatan dalam pohon maksimal dimana tiap level terdiri atas beberapa simpul. Kedalaman dihitung dari simpul utama sampai simpul terminal (simpul akhir). Pohon klasifikasi akan semakin besar jika kedalaman pohon juga semakin besar.

Tahap ketiga adalah penandaan label kelas. Pemberian label kelas untuk setiap simpul terminal berdasarkan rumus pada persamaan (5). Perbedaan warna pada tiap simpul terminal menunjukkan adanya perbedaan label kelas. Simpul terminal dengan warna biru menunjukkan pada simpul tersebut ditandai dengan label kelas 1 yang berarti anggota rumah tangga terjangkau demam tifoid, dengan persentase jumlah pengamatan yang terjangkau demam tifoid mendekati 100%. Warna biru akan berubah secara perlahan menjadi warna putih jika persentase jumlah pengamatan yang terjangkau demam tifoid pada simpul terminal tersebut berkisar 50%. Sedangkan untuk simpul terminal berwarna merah menunjukkan label kelas 2 yang berarti anggota rumah tangga tidak

terjangkit demam tifoid, dimana persentase jumlah pengamatan kelas yang tidak terjangkit

demam tifoid pada simpul tersebut mendekati 100%.



Gambar 2. Pohon Klasifikasi Maksimal

Tabel 2. Kesalahan Klasifikasi Data Learning pada Pohon Maksimal

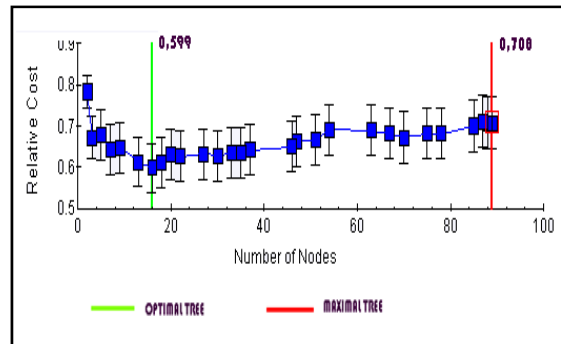
Kelas Aktual	Kelas Prediksi 1	Kelas Prediksi 2	Total Aktual
1	68	0	68
2	167	1.127	1.294
Total Prediksi	235	1.127	1.362
Benar	1	0,871	
Total Benar	0,877		

Tabel 2 menunjukkan hasil klasifikasi pohon maksimal untuk data learning. Kesalahan klasifikasi terjadi bila data pada kelas aktual 1 (terjangkit demam tifoid) masuk ke dalam kelas prediksi 2 (tidak terjangkit demam tifoid) begitupun sebaliknya. Tidak terjadi kesalahan pengklasifikasian pada kelas 1 yang merupakan kelas bagi anggota rumah tangga yang terjangkit demam tifoid. Pada kelas 2 (kelas bagi anggota rumah tangga yang tidak terjangkit demam tifoid) terjadi kesalahan pengklasifikasian sebanyak 167 pengamatan. Ketepatan klasifikasi untuk data learning pada pohon klasifikasi maksimal adalah sebesar

$$\frac{68 + 1.127}{1.362} \times 100\% = 87,7\%$$

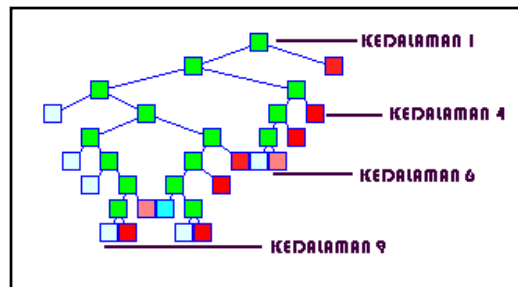
Selanjutnya dilakukan pemangkasan pohon klasifikasi maksimal. Breiman, *et al* (1993) menyatakan pemangkasan pohon klasifikasi dilakukan apabila pohon klasifikasi yang terbentuk berukuran sangat besar dan kompleks dalam penggambaran struktur data. Sehingga pada akhirnya diperoleh ukuran pohon yang layak dan berdasarkan *cost complexity minimum*.

Gambar 3 memberikan informasi bahwa nilai *relative cost* pohon klasifikasi maksimal lebih besar dibandingkan *relative cost* pohon klasifikasi optimal. Oleh karena itu perlu dilakukan pemangkasan pohon maksimal agar didapatkan nilai *relative cost* yang paling kecil. Garis hijau menunjukkan nilai *relative cost minimum* pada pohon optimal sebesar 0,599 (Persamaan 10).

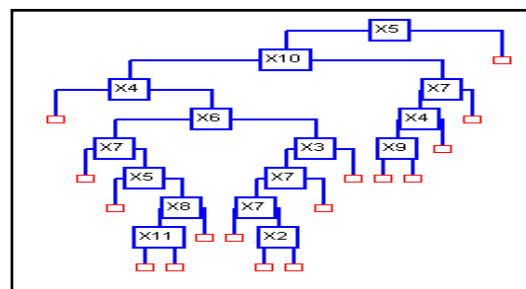


Gambar 3. Plot Relative Cost

Setelah dilakukan pemangkasan terhadap pohon klasifikasi maksimal maka dihasilkan pohon klasifikasi optimal yang memiliki relative costi terkecil dengan 9 kedalaman dan 16 simpul terminal yang disajikan dalam Gambar 4 dan *Spilters* pada pohon klasifikasi optimal disajikan pada Gambar 5.



Gambar 4. Pohon Klasifikasi Optimal



Gambar 5. Spilters Pohon Klasifikasi Optimal

Variabel prediktor yang menjadi pemilah utama pada pohon klasifikasi optimal adalah

tempat penampungan air minum (x_5) dengan skor variabel penting 100. Dengan kata lain penampungan air minum merupakan faktor utama yang mempengaruhi anggota rumah tangga terjangkau atau tidak terjangkau demam tifoid. Keterangan dari dr. Satinta Febrianti yang berdinasi di Rumah Sakit Yasmin Banyuwangi, penyebab seseorang terjangkau demam tifoid adalah bakteri *salmonella thypi*. Penularannya melalui makanan dan minuman yang telah tercemari oleh bakteri *salmonella thypi*. Orang yang kelelahan lebih mudah terjangkau penyakit demam tifoid karena daya tahan tubuhnya menurun. Apabila seseorang dengan daya tahan tubuh menurun mengkonsumsi makanan atau minuman yang tercemari oleh bakteri *s.thypi* maka orang tersebut mudah terjangkau penyakit demam tifoid. Hal ini sesuai dengan hasil penelitian ini yang menghasilkan tempat penampungan air minum sebagai faktor utama yang mempengaruhi terjangkitnya demam tifoid. Karena dengan tidak mempunyai tempat penampungan air minum atau tempat penampungan air minum terbuka maka mudah sekali bakteri *salmonella thypi* mencemari air yang merupakan bahan pokok untuk keperluan sehari-hari. Sehingga orang yang tidak mempunyai tempat penampungan air minum atau tempat penampungan air minumannya terbuka lebih rentan terjangkau demam tifoid. Selain tempat penampungan air minum variabel yang juga berkontribusi dalam pembentukan pohon optimal adalah variabel tempat penampungan air limbah (x_7) dengan skor 70.61, variabel kualitas fisik air minum (x_4) dengan skor 55.23, variabel kebiasaan cuci tangan pakai sabun sebelum makan (x_{10}) dengan skor 48.12, dan variabel tempat buang air besar (x_8) dengan skor 40.60. Variabel tempat pembuangan sampah (x_6), variabel jenis kelamin (x_2), dan variabel status sosial ekonomi (x_3) juga berkontribusi dalam pembentukan pohon optimal dengan skor variabel penting masing-masing adalah 37.50, 33.80, 22.09. Sedangkan variabel kebiasaan cuci tangan pakai sabun setelah buang air besar (x_9) dan penyuluhan kesehatan (x_{11}) memiliki skor variabel penting dibawah 20.

Simpul utama (simpul 1) dipilah oleh variabel penampungan air minum dengan mengelompokkan 931 anggota rumah tangga yang tidak memiliki tempat penampungan air minum dan anggota rumah tangga yang penampungan air minumannya terbuka pada simpul kiri menjadi simpul 2. Sisannya yaitu 431 anggota rumah tangga yang tempat penampungan air minumannya terbuka dikelompokkan pada simpul kanan menjadi simpul terminal 16. Simpul 2 terdapat 62 anggota

rumah tangga yang terjangkau demam tifoid (6,7%) dan 869 anggota rumah tangga yang tidak terjangkau demam tifoid (93,3%). Sedangkan simpul terminal 16 terdapat 6 anggota rumah tangga yang terjangkau demam tifoid (1,4%) dan 425 anggota rumah tangga yang tidak terjangkau demam tifoid (98,6%). Karena proporsi terbesar pada simpul terminal 16 adalah tidak terjangkau demam tifoid, maka pada simpul terminal 6 diberi label kelas tidak terjangkau demam tifoid (Persamaan 5). Terjadi kesalahan pengklasifikasian pada simpul terminal 16 dengan label kelas tidak terjangkau demam tifoid, karena terdapat 6 anggota rumah tangga yang dinyatakan terjangkau demam tifoid. Proses pemilahan akan terjadi lagi pada simpul 2 namun pada simpul terminal 16 tidak akan terjadi pemilahan. Simpul 2 dipilah variabel kebiasaan cuci tangan pakai sabun sebelum makan. Sebanyak 663 anggota rumah tangga yang mencuci tangan pakai sabun sebelum makan dipilah pada simpul kiri menjadi simpul 3 dan 268 anggota rumah tangga yang tidak mencuci tangan pakai sabun sebelum makan dipilah pada simpul kanan menjadi simpul 13. Pada simpul 3 terdapat 56 anggota rumah tangga yang dinyatakan terjangkau demam tifoid (8,4%) dan 607 anggota rumah tangga yang tidak terjangkau demam tifoid (91,6%). Sedangkan pada simpul 13 terdapat 6 anggota rumah tangga yang terjangkau demam tifoid (2,2%) dan 262 anggota rumah tangga yang tidak terjangkau demam tifoid (97,8). Pemilahan akan dilakukan terus-menerus sampai simpul terminal.

Tabel 3 menunjukkan hasil klasifikasi pohon maksimal untuk data learning. Kesalahan klasifikasi terjadi bila data pada kelas aktual 1 (terjangkit demam tifoid) masuk ke dalam kelas prediksi 2 (tidak terjangkau demam tifoid) begitupun sebaliknya. Jumlah kesalahan pengklasifikasian untuk kelas 1 (terjangkit demam tifoid) adalah sebanyak 17 dari 68 jumlah amatan. Jumlah kesalahan pengklasifikasian untuk kelas 2 (tidak terjangkau demam tifoid) adalah sebanyak 199 dari 1.294 jumlah amatan. Dengan demikian diperoleh ketepatan pengklasifikasian sebesar

$$\frac{51 + 1.095}{1.362} \times 100\% = 84,1\%$$

Tabel 3. Ketepatan pohon klasifikasi optimal dari data learning

Kelas Aktual	Prediksi Kelas		Total Aktual
	1	2	
1	51	17	68
2	199	1.095	1.294
Total Prediksi	250	1.112	1.362
Benar	0,750	0,846	
Total Benar	0,841		

Selanjutnya dilakukan uji validasi. Tujuan dilakukan validasi adalah untuk mengetahui layak atau tidak model pohon klasifikasi dalam pengklasifikasian data baru. Caranya yaitu data *testing* dimasukkan kedalam model pohon klasifikasi yang telah terbentuk sebelumnya dari data *learning*. Data *testing* yang digunakan sebesar 25% dari total data keseluruhan yaitu 454 data.

Tabel 4 menunjukkan bahwa data *testing* sebanyak 454 pengamatan menghasilkan ketepatan pengklasifikasian sebesar

$$\frac{19+370}{454} \times 100\% = 85,7\%.$$

Jumlah kesalahan pengklasifikasian untuk kelas 1 (terjangkit demam tifoid) adalah sebanyak 18 dari 37 jumlah amatan. Sedangkan jumlah kesalahan pengklasifikasian untuk kelas 2 (tidak terjangkit demam tifoid) adalah sebanyak 47 dari 417 jumlah amatan. Karena ketepatan klasifikasi pada data *testing* sudah tinggi yaitu 85,7% maka model pohon klasifikasi optimal yang dihasilkan sudah baik.

Tabel 4. Ketepatan pohon klasifikasi optimal dari data *testing*

Kelas Aktual	Prediksi Kelas		Total Aktual
	1	2	
1	19	18	37
2	47	370	417
Total Prediksi	66	388	454
Benar	0,514	0,887	
Total Benar	0,857		

PENUTUP

Metode *CART* menghasilkan pohon klasifikasi optimal dengan ketepatan klasifikasi data *learning* sebesar 84,1%, sedangkan ketepatan klasifikasi data *testing* adalah sebesar 85,7%. Variabel yang berpengaruh terhadap terjangkitnya penyakit demam tifoid di Aceh Utara pada pohon optimal adalah variabel tempat penampungan air minum sebagai faktor utama dengan skor tertinggi sebesar 100, tempat penampungan air limbah dengan skor 70.61, kualitas fisik air minum dengan skor 55.23, kebiasaan cuci tangan pakai sabun sebelum makan dengan skor 48.12, variabel tempat buang air besar dengan skor 40.60, tempat pembuangan sampah dengan skor 37.50, jenis kelamin dengan skor 33.80 dan status sosial ekonomi dengan skor

22.09. Sedangkan variabel kebiasaan cuci tangan pakai sabun setelah buang air besar dan penyuluhan kesehatan dengan skor variabel penting dibawah 20.

DAFTAR PUSTAKA

- [1] Anonim. 2007. *Demam Tifoid*, <<http://ummusalma.wordpress.com/2007/01/22/helloworld/>>, tanggal akses: 27 September 2009>.
- [2] Breiman L, Friedman J.H, Olshen R.A, dan Stone C.J. 1993. *Classification And Regression Trees*. Chapman And Hall. New York.
- [3] Departemen Kesehatan RI. 2008. *Riset Kesehatan Dasar (Laporan Nasional 2007)*. Jakarta.
- [4] Hidayati, N. 2001. "Analisis Regresi Poisson Terhadap Faktor-Faktor yang Mempengaruhi Penyakit Demam Typhoid di Provinsi Jawa Timur". **Skripsi, Mahasiswa Jurusan Statistika FMIPA ITS**. Surabaya.
- [5] Jevuska. 2008. **Demam Tifoid (Typhoid Fever)**, <<http://www.jevuska.com/2008/05/10-/demam-tifoid-typhoid-fever>>, tanggal akses: 26 September 2009>.
- [6] Kompas. 2005. **Masyarakat Diminta Waspada Penyakit Tipus**, <<http://www.kompas.com/kompas-cetak/>>, tanggal akses: 28 Agustus 2009>.
- [7] Lewis dan Roger J. 2000. **An Introduction to Classification And Regression Trees (CART) Analysis**. Presented at the 2000.
- [8] Lubis, R. (2007). "Faktor Resiko Kejadian Penyakit Demam Tifoid Penderita Yang Dirawat Di RSUD DR. Soetomo Surabaya". **Thesis, Mahasiswa Jurusan Ilmu Kesehatan Masyarakat Universitas Sumatera Utara**. Sumatera Utara.
- [9] Salma, U. 2007. **Demam Tifoid**, <<http://ummusalma.word-press.com>>, tanggal akses 23 Agustus 2009>.
- [10] Steinberg D. dan Phillip C. 2005. **CART-Classification and Regression Trees**. CA: Salford System, San Diego.