

# Klasifikasi Risiko Infeksi pada Bayi Baru Lahir di Rumah Sakit Umum Daerah Sidoarjo Menggunakan Metode *Classification Trees*

Aulia Rahma Safitri dan Sri Pingit Wulandari  
Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Institut Teknologi Sepuluh Nopember (ITS)  
Jl.Arief Rahman Hakim, Surabaya 60111 Indonesia  
*e-mail:* sri\_pingit@statistika.its.ac.id

**Abstrak**—Angka kematian bayi (AKB) merupakan salah satu indikator derajat kesehatan masyarakat yang dapat mengukur tingkat kesejahteraan masyarakat ataupun tingkat kemiskinan di Indonesia. AKB pada tahun 2012 sebesar 32 per 1.000 kelahiran hidup dan proporsi kematian neonatal terhadap kematian bayi meningkat menjadi 59,4%. Infeksi pada bayi baru lahir gejalanya kurang jelas dan seringkali tidak diketahui sampai keadaannya sudah sangat terlambat. Informasi tentang variabel yang paling berpengaruh terhadap risiko infeksi pada bayi akan membantu untuk menurunkan kasus infeksi. Selain itu, karakteristik bayi baru lahir dapat diketahui guna membantu menetapkan status risiko infeksi, sehingga bayi dengan status positif risiko infeksi dapat diberikan perawatan intensif. Penelitian ini berguna untuk mengetahui hasil ketepatan klasifikasi yang sudah dilakukan oleh RSUD Sidoarjo dan juga untuk mengetahui variabel terpenting yang berguna untuk pengklasifikasian. Sebelum data digunakan dalam analisis *Classification Trees*, dilakukan *pre-processing* untuk mengatasi *missing value*, dengan menghilangkan data yang mengandung *missing*. Berdasarkan hasil analisis *Classification Trees* dengan aturan pemilihan pemilah indeks gini, diperoleh akurasi untuk data *testing* sebesar 93,5%, dengan persentase *sensitivity* dan *specificity* masing-masing sebesar 93,1% dan 93,9%. Faktor terpenting dalam klasifikasi risiko infeksi pada bayi baru lahir di RSUD Sidoarjo adalah kondisi sisa air ketuban.

**Kata Kunci**—Akurasi, Angka Kematian Bayi, *Classification Trees*, Indeks Gini, Risiko Infeksi.

## I. PENDAHULUAN

Angka kematian bayi (AKB) merupakan salah satu indikator derajat kesehatan masyarakat yang termasuk dalam salah satu tujuan dari Rencana Strategis Kementerian Kesehatan Republik Indonesia untuk tahun 2015-2019, yaitu menurunkan AKB menjadi 24 per 1.000 kelahiran hidup. AKB berhasil diturunkan secara tajam menjadi 32 per 1.000 kelahiran hidup pada 2012 [1]. Kontribusi terbesar terhadap kematian bayi diberikan oleh kematian neonatal. Kematian neonatal adalah kematian bayi dalam bulan pertama setelah lahir. Proporsi kematian neonatal terhadap kematian bayi meningkat dari 47,1% pada 2009 menjadi 59,4% pada 2012. Berdasarkan hasil SUSENAS Jawa Timur pada 2013, AKB Kabupaten Sidoarjo sebesar 22,11 per 1.000 kelahiran hidup [2]. Menurut Dinas Kesehatan Kabupaten Sidoarjo penyebab utama kematian bayi adalah bayi berat lahir rendah, prematuritas dan infeksi.

Kasus infeksi merupakan salah satu penyakit yang dialami bayi baru lahir. Umumnya, kasus infeksi menunjukkan gejala yang kurang jelas dan seringkali tidak diketahui sampai keadaannya sudah sangat terlambat. Jika faktor penyebab risiko infeksi yang paling dominan dapat diketahui, maka hal tersebut akan membantu untuk menurunkan kasus infeksi. Selain itu, diagnosa yang tepat dari tenaga kesehatan untuk menetapkan status risiko infeksi pada bayi baru lahir merupakan hal yang penting, sehingga bayi yang berisiko infeksi dapat diberikan perawatan intensif dan beberapa antibiotik untuk melawan bakteri yang menyebabkan infeksi. Karakteristik bayi baru lahir dapat diketahui melalui pemeriksaan esensial berdasarkan kondisi bayi sesaat setelah lahir untuk membedakan status risiko infeksi pada bayi baru lahir. Hal tersebut dapat diwujudkan dengan melakukan pengklasifikasian pada bayi baru lahir. Metode klasifikasi yang dapat digunakan adalah metode *Classification Trees*, yang merupakan bagian dari metode CART (*Classification and Regression Trees*). Metode CART adalah metode nonparametrik yang akan menghasilkan pohon klasifikasi apabila variabel respon yang digunakan bersifat kategori dan akan menghasilkan pohon regresi jika variabel respon yang digunakan bersifat kontinyu [3].

Menelaah dari pemikiran tersebut, maka dalam penelitian ini dilakukan klasifikasi menggunakan metode *Classification Trees* untuk mengetahui ketepatan klasifikasi dan faktor-faktor yang mempengaruhi pengklasifikasian bayi baru lahir berdasarkan status risiko infeksi. Studi kasus yang digunakan dalam penelitian adalah bayi baru lahir di RSUD Sidoarjo. Karena rumah sakit yang menjadi tujuan utama bagi ibu hamil untuk melakukan persalinan di Kabupaten Sidoarjo adalah RSUD Sidoarjo. Berdasarkan rekam medis RSUD Sidoarjo, kejadian kasus infeksi di rumah sakit tersebut menunjukkan angka yang cukup tinggi, yakni sebanyak 17 bayi positif risiko infeksi dari total 35 bayi yang lahir pada awal Januari 2015. Hasil akurasi dan karakteristik yang dihasilkan dapat dijadikan sebagai informasi kepada Dinas Kesehatan atau rumah sakit setempat, guna menurunkan kasus infeksi pada bayi.

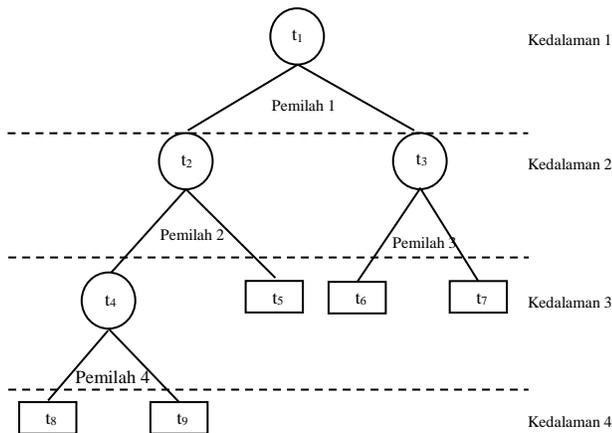
## II. TINJAUAN PUSTAKA

### A. *Classification and Regression Trees* (CART)

Metode CART digunakan untuk menggambarkan hubungan antara variabel respon dengan satu atau lebih variabel pre-

diktor. Analisis CART memiliki sifat “*binary recursive partitioning*”. Istilah “*binary*” mengimplikasikan bahwa sekelompok data yang terkumpul dalam suatu ruang direpresentasikan sebagai simpul atau *node* pada pohon keputusan, dimana *node* tersebut hanya dapat diklasifikasikan menjadi dua simpul anak (*child nodes*), dan simpul utama disebut sebagai *parent node*. Istilah “*recursive*” dihubungkan pada kenyataan bahwa proses penyekatan secara biner dilakukan secara berulang-ulang, sedangkan istilah “*partitioning*” memiliki arti bahwa proses klasifikasi dapat dilakukan dengan cara memilah kumpulan data menjadi beberapa bagian atau partisi [4].

Ilustrasi struktur pohon klasifikasi yang ditunjukkan pada Gambar 1 [3]. Simpul awal yang mengandung seluruh data dengan notasi  $t_1$ . Simpul dalam (*internal node*) dinotasikan dengan  $t_2, t_3$ , dan  $t_4$ , sedangkan simpul akhir (*terminal node*) dinotasikan dengan  $t_5, t_6, t_7, t_8$ , dan  $t_9$  dimana setelahnya tidak ada lagi pemilahan, artinya simpul anak yang dihasilkan telah homo-gen. Setiap simpul berada pada kedalaman (*depth*) tertentu, dimulai dari simpul awal  $t_1$  yang berada pada kedalaman 1,  $t_2$  dan  $t_3$  berada pada kedalaman 2, dan begitu seterusnya hingga didapatkan simpul  $t_4, t_5, t_6, t_7, t_8$ , dan  $t_9$  yang berada pada kedalaman 4.



Gambar 1. Ilustrasi Pohon Klasifikasi

Analisis CART memiliki tiga tahap dasar, yakni pembentukan pohon klasifikasi maksimal, pemangkasan dan penentuan pohon klasifikasi optimal.

### 1. Pembentukan Pohon Klasifikasi Maksimal

Pada pembentukan pohon klasifikasi diawali dengan menentukan variabel apa dan nilai mana dari variabel tersebut yang layak dijadikan pemilah bagi setiap simpul. Dalam pembentukan pohon klasifikasi dibutuhkan data *learning L* yang terdiri atas pengamatan berukuran  $N$ . Proses pembentukan pohon klasifikasi terdiri dari tiga tahap, yaitu pemilihan pemilah, penentuan simpul terminal, dan penandaan label kelas.

#### a. Pemilihan Pemilah

Pemilihan pemilah bertujuan untuk mengurangi tingkat keheterogenan pada simpul induk dan mendapatkan simpul anak dengan tingkat homogenitas yang tinggi. Fungsi keheterogenan yang digunakan adalah indeks gini yang dituliskan dalam persamaan sebagai berikut.

$$i(t) = \sum_{i,j=1} p(j|t)p(i|t), i \neq j \tag{1}$$

Nilai  $i(t)$  menunjukkan *impurity* atau fungsi keheterogenan simpul  $t$ ,  $p(j|t)$  adalah proporsi kelas  $j$  pada simpul  $t$ , dan  $p(i|t)$

adalah proporsi kelas  $i$  pada simpul  $t$ . Kemudian pemilihan pemilah terbaik dilakukan berdasarkan kriteria *goodness of split* ( $\phi(s,t)$ ) yang didefinisikan sebagai penurunan heterogenitas dengan formula sebagai berikut.

$$\phi(s,t) = \Delta i(s,t) = i(t) - p_L i(t_L) - p_R i(t_R) \tag{2}$$

Dimana  $\phi(s,t)$  adalah nilai *goodness of split*,  $i(t)$  adalah fungsi heterogenitas pada simpul  $t$ ,  $P_L$  adalah proporsi pengamatan simpul kiri,  $P_R$  adalah proporsi pengamatan menuju simpul kanan,  $i(t_L)$  adalah fungsi heterogenitas pada simpul anak kiri, dan  $i(t_R)$  adalah fungsi heterogenitas pada simpul anak kanan.

#### b. Penentuan Simpul Terminal

Penentuan simpul terminal merupakan tahap kedua dimana simpul  $t$  akan diputuskan untuk dipilah lagi atau tidak dipilah lagi apabila terdapat penurunan heterogenitas yang signifikan.

#### c. Penandaan Label Kelas

Penandaan label kelas pada simpul terminal ditentukan berdasarkan aturan jumlah terbanyak, yaitu apabila,

$$p(j_0|t) = \max_j p(j|t) = \max_j \frac{N_j(t)}{N(t)} \tag{3}$$

Dengan  $p(j|t)$  adalah proporsi kelas  $j$  pada simpul  $t$ ,  $N_j(t)$  adalah banyaknya amatan kelas  $j$  pada *terminal node*,  $N(t)$  adalah jumlah total pengamatan dalam *terminal node*. Label kelas untuk *terminal node* adalah  $j_0$ . Pemberian label kelas  $j_0$  tersebut memberikan nilai dugaan kesalahan pengklasifikasian pada simpul  $t$  paling kecil yaitu sebesar  $r(t) = 1 - \max_j p(j|t)$ .

### 2. Pemangkasan Pohon Klasifikasi

Untuk menghindari kasus *overfitting* dan *underfitting* dari pohon klasifikasi maksimal, makaperlu dilakukan pemangkasan pohon (*pruning*) dengan fungsi *cost complexity* adalah sebagai berikut [3].

$$R_\alpha(T) = R(T) + \alpha |\tilde{T}| \tag{4}$$

$R_\alpha(T)$  adalah *cost complexity measure* (ukuran kompleksitas suatu pohon  $T$  pada kompleksitas  $\alpha$ ),  $R(T)$  adalah penduga pengganti (*resubstitution estimate*) pohon atau ukuran kesalahan klasifikasi pohon  $T$ ,  $\alpha$  adalah *complexity parameter* (*cost* bagi penambahan satu simpul akhir pada pohon  $T$ ), dan  $|\tilde{T}|$  adalah ukuran banyaknya *terminal node* pada pohon  $T$ . Bagian pohon  $T$  yang dapat meminimumkan  $R_\alpha(T)$  merupakan bagian pohon yang akan dipangkas.

### 3. Penentuan Pohon Klasifikasi Optimal

Penduga yang dapat digunakan untuk mendapatkan pohon klasifikasi optimal adalah penduga *v-fold cross validation estimate*. Nilai  $v$  yang sering digunakan adalah 10, sehingga menjadi *10-fold cross validation estimate* [5]. Penduga sampel uji untuk  $T_i^{(v)}$  adalah sebagai berikut.

$$R(T_i^{(v)}) = \frac{1}{N} \sum_{(x_n, j_n) \in L_v} X(d^{(v)}(x_n) \neq j_n) \tag{5}$$

Dimana,  $d^{(v)}(x)$  adalah hasil pengklasifikasian,  $N_v$  adalah jumlah pengamatan dalam  $L_v$ . Penduga validasi *v-fold cross validation* untuk  $T_i^{(v)}$  adalah sebagai berikut.

$$R^{cv}(T_i) = \frac{1}{V} \sum_{v=1}^V R^{cv}(T_i^{(v)}) \tag{6}$$

Sehingga, pohon klasifikasi yang optimum dipilih  $T^*$  dengan  $R^{cv}(T^*) = \min R^{cv}(T_i)$ .

4. Ukuran Ketepatan Klasifikasi

Untuk mengevaluasi hasil ketepatan klasifikasi yaitu dengan cara menghitung akurasi klasifikasi. Pada penelitian ini evaluasi akurasi klasifikasi ialah dengan menghitung nilai *total accuracy rate* (1-APER). Selain itu terdapat ukuran *sensitivity* yang menggambarkan akurasi pada sampel kelas  $i$ , sedangkan *specificity* menggambarkan akurasi pada kelas  $j$ . Metode klasifikasi yang baik seharusnya mampu mengukur *sensitivity* dan *specificity* sama baiknya. *Crosstab* untuk menghitung ketepatan klasifikasi ditunjukkan dalam Tabel 1 sebagai berikut.

Tabel 1.  
Tabulasi Silang Hasil Ketepatan Klasifikasi

Kelas Pengamatan Y	Kelas Prediksi Y		Total
	1	2	
1	$n_{11}$	$n_{12}$	$N_1$
2	$n_{21}$	$n_{22}$	$N_2$
Total	$N_1$	$N_2$	$N$

Berikut ini adalah formula untuk menghitung total akurasi, *sensitivity* dan *specificity*.

$$\text{Total Accuracy Rate (1-APER)} = \frac{n_{11} + n_{22}}{N} \tag{7}$$

$$\text{Sensitivity} = \frac{n_{11}}{N_1} \tag{8}$$

$$\text{Specificity} = \frac{n_{22}}{N_2} \tag{9}$$

B. Periode Perinatal dan Neonatal Dini

Bayi yang termasuk dalam periode perinatal dimulai dari saat bayi masih dalam kandungan dengan usia kehamilan 22 minggu atau ketika bayi lahir dengan berat lahir 500 gram, sampai pada bayi berusia 7 hari. Sedangkan bayi yang berusia kurang dari sama dengan 7 hari termasuk periode neonatal dini [6].

C. Infection Specific to the Perinatal Period

Umumnya, infeksi khusus yang terjadi pada periode perinatal dianggap kongenital dan dapat diklasifikasikan dalam diagnosa perinatal dengan aturan kode internasional ICD-10 *medical condition (International Classification of Disease)*, dimana kode diagnosanya yaitu P35-P39. Bayi diberikan diagnosa tambahan dengan kode P35-P39 dapat diperoleh sebelum kelahiran melalui umbilikus atau setelah proses kelahiran. Infeksi yang terjadi setelah lahir, tetapi muncul selama periode neonatal dini dapat dilihat apakah ada indikasi terkena paparan infeksi bukan dari bawaan lahir. Oleh sebab itu, dokter harus memberikan klarifikasi mengenai diagnosa infeksi pada bayi, ketika catatan rekam medis tidak sepenuhnya jelas. Untuk kode P39.9 merupakan kode diagnosa yang diberikan jika bayi belum diketahui infeksi spesifiknya atau dalam dunia kesehatan biasa disebut *infection unspecified to the perinatal period*[7].

D. Perawatan Neonatal Esensial Pada Saat Lahir

Bayi baru lahir sangat rentan terhadap infeksi yang disebabkan oleh paparan atau kontaminasi mikroorganisme selama proses persalinan berlangsung maupun beberapa saat

setelah lahir. Setelah bayi lahir, bayi diletakkan di atas kain bersih dan kering yang telah disiapkan pada perut bawah ibu. Kemudian dilakukan penilaian lagi mengenai tangisan bayi, pernafasan bayi, warna kulit, denyut jantung, tonus otot bayi baik atau bayi bergerak aktif, pengukuran suhu ketiak, meraba bagian kepala, mata, mulut, perut, tali pusat, punggung atau tulang belakang, lubang anus, alat kelamin luar, ekstremitas (jumlah jari tangan dan kaki), timbang bayi, mengukur panjang dan lingkaran kepala bayi dan menilai cara menyusui. Jika bayi baru lahir cukup bulan dengan air ketuban jernih dan langsung menangis atau bernapas spontan serta bergerak aktif cukup dilakukan perawatan bayi baru lahir normal [8].

Tanda atau gejala bayi termasuk klasifikasi infeksi bakteri berat antara lain tidak mau minum, riwayat kejang, bergerak hanya jika dirangsang, napas cepat ( $\geq 60$  kali/menit), napas lambat ( $<30$  kali/menit), tarikan dinding dada ke dalam yang kuat, merintih, demam ( $> 37,5^\circ\text{C}$ ), hipotermia ( $< 35,5^\circ\text{C}$ ), nanah di mata atau pusar kemerahan meluas sampai dinding perut. Jika bayi termasuk pada klasifikasi penyakit sangat berat atau infeksi berat, maka bayi perlu diberikan antibiotik intramuskular. Antibiotik pilihan pertama adalah ampisilin dan gentamisin.

III. METODOLOGI PENELITIAN

A. Sumber Data

Data yang digunakan dalam penelitian tugas akhir ini merupakan data sekunder yang diambil dari data rekam medis bayi baru lahir di RSUD Sidoarjo pada bulan September–Oktober 2015. Jumlah bayi yang lahir di RSUD Sidoarjo dalam periode waktu tersebut sebanyak 222 bayi dan bayi baru lahir yang dirujuk ke RSUD Sidoarjo dalam periode waktu yang sama sebanyak 25 bayi. Total data rekam medis yang dikumpulkan dalam penelitian ini sebanyak 247 bayi.

B. Variabel Penelitian

Variabel yang digunakan dalam penelitian ini adalah variabel respon dan variabel prediktor. Variabel respon yang digunakan adalah status risiko infeksi pada bayi dengan kategori risiko infeksi negatif dan positif, sedangkan variabel prediktor yang digunakan meliputi 10 variabel yang ditunjukkan dalam Tabel 2 sebagai berikut.

Tabel 2.  
Variabel Penelitian

Variabel	Nama Variabel	Kategori	Skala Pengukuran
Y	Status Risiko Infeksi	0: Negatif 1: Positif	Nominal
$X_1$	Berat Badan Lahir	-	Rasio
$X_2$	Bayi Lahir Prematur	0: Tidak 1: Ya	Nominal
$X_3$	Kondisi Air Ketuban	0: Normal 1: Keruh/Hijau	Nominal
$X_4$	Ketuban Pecah Dini (KPD)	0: Tidak 1: Ya	Nominal
$X_5$	Suhu Tubuh	-	Interval
$X_6$	Warna Kulit	0: Blue 1: All Pink	Nominal
$X_7$	Denyut Jantung	-	Rasio
$X_8$	Refleks Rangsangan	0: Menangis keras 1: Merintih	Nominal
$X_9$	Kecepatan Bernafas	0: Teratur 1: Tidak teratur	Nominal
$X_{10}$	Kondisi Tali Pusat	0: Segar	Nominal

C. Langkah Analisis

Langkah analisis yang dilakukan dalam penelitian menggunakan metode *classification trees* adalah sebagai berikut.

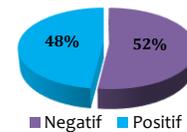
1. Melakukan *pre-processing* terhadap data yang telah dikumpulkan untuk meningkatkan akurasi pada hasil klasifikasi dengan mengatasi kasus *missing value*.
2. Mendiskripsikan karakteristik bayi baru lahir berdasarkan dua tipe variabel respon menggunakan nilai rata-rata, deviasi standar, maksimum, minimum, *pie-chart* dan *bar chart*.
3. Membagi data secara random sebanyak 10 bagian berdasarkan aturan *10-fold cross validation estimate*, dimana 9 bagian sebagai data *learning* dan 1 bagian sebagai data *testing*.
4. Melakukan analisis klasifikasi dengan metode *classification trees* melalui tahapan sebagai berikut.
  - a. Membentuk pohon klasifikasi maksimal menggunakan data *learning* dengan tahapan sebagai berikut.
    - 1) Melakukan pemilihan pemilah berdasarkan variabel prediktor menurut aturan indeks gini dan kriteria *goodness of split*.
    - 2) Menentukan simpul terminal.
    - 3) Melakukan penandaan label kelas pada simpul terminal berdasarkan jumlah terbanyak dari tiap kelas yang ada pada variabel respon.
  - b. Memangkas pohon klasifikasi berdasarkan kriteria *cost complexity minimum*.
  - c. Menentukan pohon klasifikasi yang optimal.
  - d. Menghitung ketepatan klasifikasi dan melakukan validasi pohon hasil bentukan menggunakan data *testing*. Ukuran ketepatan klasifikasi yang digunakan adalah *total accuracy rate*, *sensitivity* dan *specificity*.
5. Menarik kesimpulan mengenai perbandingan karakteristik bayi baru lahir berisiko infeksi dan bayi baru lahir tidak berisiko infeksi.

IV. HASIL DAN PEMBAHASAN

A. Karakteristik Bayi Baru Lahir di RSUD Sidoarjo

Jumlah bayi yang dikumpulkan melalui catatan rekam medis RSUD Sidoarjo sebanyak 247 bayi. Berdasarkan jumlah tersebut sebanyak 130 bayi lahir normal atau memiliki status negatif risiko infeksi, sedangkan 117 bayi lahir dengan status positif risiko infeksi. Berdasarkan indikator yang diperoleh terdapat beberapa kasus *missing value* pada dua indikator, yakni suhu tubuh bayi ( $X_5$ ) dan denyut jantung ( $X_7$ ) dengan jumlah kasus *missing value* masing-masing sebanyak 1 dan 6 kasus. Karena jumlah *missing value* yang terdapat dalam indikator sebanyak 2,43% (kurang dari 30%), maka diputuskan untuk menghapus data yang mengandung *missing value* guna memberikan informasi yang lebih akurat dan dapat menghasilkan performansi yang lebih baik. Data yang dihapuskan sebanyak 7 data, yakni 5 data termasuk dalam kategori status negatif dan 2 data termasuk dalam kategori status positif berisiko infeksi, sehingga jumlah data yang tersisa sebanyak 240 dengan 125 bayi negatif risiko infeksi

dan sebanyak 115 bayi positif risiko infeksi. Deskripsi status risiko infeksi pada bayi baru lahir di RSUD Sidoarjo disajikan pada Gambar 2 sebagai berikut.



Gambar 2. Deskripsi Status Risiko Infeksi pada Bayi

B. Analisis *Classification Trees* untuk Klasifikasi Risiko Infeksi pada Bayi Baru Lahir di RSUD Sidoarjo

Jumlah data yang digunakan dalam analisis ini sebanyak 200 data, sedangkan 40 data digunakan sebagai data bayi yang baru lahir. Pembentukan pohon klasifikasi maksimal diawali dengan pemilihan *classifier* atau pemilah, dengan memilih variabel prediktor terpenting yang akan digunakan sebagai pemilah awal. Berdasarkan variabel prediktor yang digunakan dalam penelitian ini, proses pemilihan pemilah dilakukan dengan mengacu pada skor kontribusi terbesar yang dihasilkan dari tiap variabel prediktor. Besar kontribusi yang dihasilkan oleh masing-masing variabel prediktor dapat dilihat dalam Tabel 3 sebagai berikut.

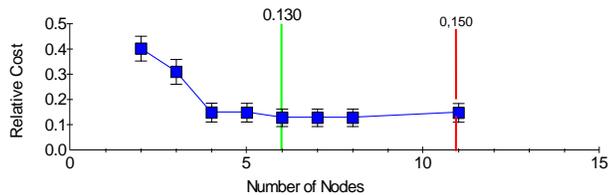
Tabel 3  
Skor Variabel Prediktor dalam Pembentukan Pohon Klasifikasi Maksimal

Variabel	Nama Variabel	Skor
$X_3$	Kondisi Air Ketuban	100,00
$X_9$	Kecepatan Bernafas	41,559
$X_8$	Refleks Rangsangan	41,599
$X_4$	Ketuban Pecah Dini	35,237
$X_1$	Berat Badan Lahir	29,357
$X_{10}$	Kondisi Tali Puser	24,050
$X_5$	Suhu Tubuh	17,775
$X_2$	Lahir Prematur	10,102
$X_7$	Denyut Jantung	9,664
$X_6$	Warna Kulit	6,582

Tabel 3 menunjukkan bahwa variabel terpenting yang memiliki kontribusi terbesar adalah variabel  $X_3$  (kondisi air ketuban) dengan skor 100. Artinya, variabel  $X_3$  akan menjadi pemilah awal atau biasa disebut sebagai simpul induk. Faktor kondisi air ketuban mampu menurunkan tingkat keheterogenan paling besar, sehingga simpul yang dihasilkan akan lebih homogen.

Selanjutnya dilakukan proses pemilahan dengan fungsi keheterogenan indeks gini hingga terbentuk pohon klasifikasi maksimal. Jumlah simpul terminal yang dihasilkan oleh pohon klasifikasi maksimal adalah sebanyak 11 simpul dengan tingkat kedalaman sebesar 7.

Setelah memperoleh pohon klasifikasi maksimal, langkah berikutnya adalah pemangkasan, yang bertujuan untuk menghindari adanya kasus *overfitting* atau *underfitting*. Proses pemangkasan pohon dilakukan berdasarkan pendekatan *cost complexity minimum*. Setiap hasil pemangkasan yang dilakukan memiliki nilai *relative cost* tertentu, kemudian dipilih hasil pemangkasan yang mempunyai nilai *relative cost* paling minimum. Plot *relative cost* dari setiap pemangkasan yang dilakukan pada pohon klasifikasi risiko infeksi pada bayi baru lahir di RSUD Sidoarjo yang ditampilkan dalam Gambar 3 sebagai berikut.



Gambar 3. Nilai *Relative Cost* Hasil Pemangkasan Pohon Klasifikasi

Gambar 3 menjelaskan bahwa pohon klasifikasi maksimal yang terbentuk perlu dilakukan pemangkasan seperti yang ditunjukkan oleh garis berwarna merah dengan nilai *relative cost* sebesar 0,150. Sedangkan garis berwarna hijau dalam Gambar 3 menunjukkan bahwa pohon klasifikasi tersebut telah optimal dengan nilai *relative cost* paling minimum yaitu sebesar 0,130. Hasil pemangkasan yang keempat tersebut menghasilkan pohon klasifikasi optimal dengan jumlah simpul terminal sebanyak 6 simpul. Urutan mulai dari pohon klasifikasi maksimal yang dibentuk hingga pemangkasan yang dilakukan terhadap pohon tersebut dapat dilihat dalam Tabel 4 berikut.

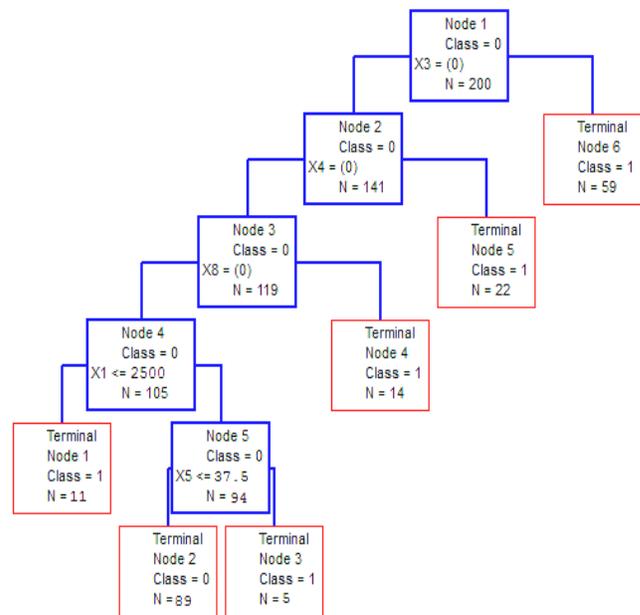
Tabel 4.  
Urutan Pohon Klasifikasi

Tree Number	Terminal Nodes	Cross-Validated Relative Cost	Resubstitution Relative Cost	Complexity Parameter
1	11	0,150 ± 0,037	0,070	0,000
4**	6	0,130 ± 0,035	0,080	0,002
5	5	0,150 ± 0,037	0,110	0,005
8	2	0,404 ± 0,049	0,404	0,071
9	1	1,000 ± 0,000	1,000	0,298

\*\*pohon klasifikasi optimal

Biaya kesalahan (*cross-validated relative cost*) yang dihasilkan oleh setiap hasil pemangkasan pohon klasifikasi maksimal. Pohon pertama merupakan pembentukan pohon klasifikasi maksimal dengan jumlah simpul terminal sebanyak 11 simpul dan biaya kesalahan berkisar antara 0,150 dan 0,037. Sedangkan pohon keempat merupakan pohon klasifikasi yang menghasilkan biaya kesalahan paling minimum yakni berkisar antara 0,130 dan 0,035. Oleh sebab itu, pohon keempat merupakan pohon klasifikasi optimal dengan jumlah simpul terminal sebanyak 6 simpul dan tingkat kedalaman sebesar 6.

Pohon klasifikasi optimal yang ditunjukkan dalam Gambar 4 dapat menjelaskan bahwa variabel yang menentukan klasifikasi risiko infeksi pada bayi baru lahir adalah variabel  $X_3$  (kondisi air ketuban). Pemilahan diawali dengan memilah data sebanyak 200 bayi berdasarkan kondisi air ketuban. Kemudian data dibagi menjadi dua bagian yaitu simpul kiri (*node 2*) yang terdiri dari 141 bayi dengan kondisi air ketuban yang jernih, dan dilabeli sebagai bayi negatif risiko infeksi. Sedangkan, simpul kanan (*terminal node 7*) terdiri dari 59 bayi dengan kondisi air ketuban yang keruh dan berwarna hijau, setelah itu diberi label kelas sebagai bayi baru lahir positif risiko infeksi. Selanjutnya, pemilahan kembali dilakukan secara berulang karena tingkat heterogenitas masih tinggi. Pemilahan berhenti dengan jumlah terminal simpul sebanyak 6 simpul.



Gambar 4. Pohon Klasifikasi Optimal

Karakteristik simpul terminal yang mengklasifikasikan data menjadi salah satu dari kedua status risiko infeksi dijelaskan sebagai berikut.

1. Simpul terminal 1 terdiri dari 11 bayi yang diprediksi berisiko infeksi. Karakteristik bayi pada simpul ini adalah memiliki berat badan lahir kurang dari sama dengan 2500 gram, refleks bayi setelah diberi rangsangan menangis keras atau terdapat gerakan melawan, tidak mengalami ketuban pecah dini dan sisa air ketuban dalam kondisi jernih.
2. Simpul terminal 2 terdiri dari 89 bayi yang diprediksi tidak memiliki risiko infeksi. Karakteristik bayi pada simpul ini antara lain adalah bayi memiliki suhu tubuh kurang dari sama dengan 37,5°C, berat badan lahir bayi lebih dari 2500 gram, refleks bayi setelah diberi rangsangan menangis keras atau terdapat gerakan melawan, tidak mengalami ketuban pecah dini dan sisa air ketuban dalam kondisi jernih.
3. Simpul terminal 3 terdiri dari 5 bayi yang diprediksi memiliki risiko infeksi. Karakteristik bayi yang terbentuk adalah suhu tubuh bayi lebih dari 37,5°C, berat badan lahir lebih dari 2500 gram, refleks bayi setelah diberi rangsangan menangis keras atau terdapat gerakan melawan, tidak mengalami ketuban pecah dini dan sisa air ketuban dalam kondisi jernih.
4. Simpul terminal 4 terdiri dari 14 bayi yang diprediksi memiliki risiko infeksi. Bayi yang termasuk dalam simpul ini memiliki beberapa karakteristik yaitu ketika bayi diberi rangsangan, bayi menunjukkan refleks merintih atau menangis lemah, tidak mengalami ketuban pecah dini dan sisa air ketuban dalam kondisi jernih.
5. Simpul terminal 5 terdiri dari 22 bayi yang diprediksi berisiko infeksi. Bayi yang termasuk dalam simpul ini memiliki beberapa karakteristik antara lain adalah ibu bayi mengalami ketuban pecah dini dan sisa air ketuban dalam kondisi jernih.
6. Simpul terminal 6 terdiri dari 59 bayi yang diprediksi memiliki risiko infeksi. Karakteristik bayi yang termasuk

simpul ini antara lain adalah sisa air ketuban dalam kondisi keruh atau berwarna hijau.

Sesuai dengan pemaparan di atas, karakteristik bayi baru lahir berdasarkan faktor-faktor yang mempengaruhi bayi berisiko infeksi ditampilkan dalam Tabel 5 sebagai berikut.

Tabel 5.

Karakteristik Bayi Baru Lahir Berdasarkan Simpul Terminal	
Negatif Risiko Infeksi	Positif Risiko Infeksi
✓ Bayi dengan suhu tubuh $\leq 37,5^{\circ}\text{C}$ , berat badan lahir $> 2500$ gram yang menangis keras saat diberi rangsangan, tidak mengalami ketuban pecah dini dan sisa air ketuban dalam kondisi jernih	✓ Bayi dengan sisa air ketuban dalam kondisi keruh ✓ Bayi dengan sisa air ketuban jernih dan mengalami ketuban pecah dini ✓ Bayi yang merintih ketika diberi rangsangan, tidak mengalami ketuban pecah dini dan kondisi sisa air ketuban jernih ✓ Bayi dengan berat badan lahir $\leq 2500$ gr yang menangis ketika diberi rangsangan, tidak mengalami ketuban pecah dini dan kondisi sisa air ketuban jernih ✓ Bayi dengan suhu tubuh $>37,5^{\circ}\text{C}$ yang memiliki berat badan lahir $> 2500$ gr, menangis ketika diberi rangsangan, tidak mengalami ketuban pecah dini dan kondisi sisa air ketuban jernih

Langkah selanjutnya adalah menghitung ketepatan klasifikasi dari pohon klasifikasi optimal, untuk mengetahui layak atau tidaknya pohon klasifikasi optimal yang terbentuk perlu dilakukan validasi menggunakan data *testing*. Hasil ketepatan klasifikasi risiko infeksi pada bayi baru lahir di RSUD Sidoarjo menggunakan data *testing* ditunjukkan dalam Tabel 6.

Tabel 6.

Aktual	Prediksi		Total
	Negatif	Positif	
Negatif	94	7	101
Positif	6	93	99
<b>Total</b>	100	100	200
<b>Total Akurasi</b>			0,935
<i>Sensitivity</i>			0,931
<i>Specificity</i>			0,939

Sebanyak 94 bayi tepat diklasifikasikan sebagai bayi baru lahir dengan status negatif risiko infeksi. Sedangkan sebanyak 93 bayi tepat diklasifikasikan sebagai bayi baru lahir positif risiko infeksi. Total akurasi (1-APER) yang dihasilkan untuk klasifikasi risiko infeksi pada bayi baru lahir di RSUD Sidoarjo sebesar 93,5%, sedangkan total kesalahan klasifikasi (APER) yang dihasilkan sebesar 6,5% dengan nilai *sensitivity* dan *specificity* masing-masing sebesar 93,1% dan 93,9%. Karena total akurasi yang cukup tinggi, maka pohon klasifikasi optimal yang terbentuk sudah layak digunakan untuk pengklasifikasian risiko infeksi pada bayi yang baru lahir berikutnya.

## V. KESIMPULAN DAN SARAN

Berdasarkan hasil analisis *Classification Trees* yang telah dilakukan untuk pengklasifikasian risiko infeksi pada bayi baru lahir di RSUD Sidoarjo diperoleh akurasi sebesar 93,5% dengan persentase *sensitivity* dan *specificity* masing-masing sebesar 93,1% dan 93,9%. Variabel yang paling dominan untuk mengklasifikasi status risiko infeksi pada bayi baru lahir adalah kondisi sisa air ketuban. Karakteristik pada bayi baru lahir dengan status negatif risiko infeksi adalah bayi suhu tubuh  $\leq 37,5^{\circ}\text{C}$ , berat badan lahir  $> 2500$  gram yang menangis keras saat diberi stimulus, tidak mengalami ketuban pecah dini dan dan sisa air ketuban dalam kondisi jernih. Sedangkan karakteristik bayi baru lahir dengan status positif risiko infeksi adalah bayi dengan sisa air ketuban dalam kondisi keruh; bayi dengan kondisi sisa air ketuban jernih dan mengalami KPD; bayi yang merintih ketika diberi stimulan, tidak mengalami KPD dan kondisi sisa air ketuban jernih; bayi dengan berat badan lahir  $\leq 2500$  gram yang menangis ketika diberi stimulan, tidak mengalami KPD dan kondisi sisa air ketuban jernih; serta bayi dengan suhu tubuh  $> 37,5^{\circ}\text{C}$  yang memiliki berat badan lahir  $> 2500$  gram, menangis ketika diberi stimulan, tidak mengalami KPD dan kondisi sisa air ketuban jernih.

Rekomendasi yang dapat diberikan untuk penelitian selanjutnya adalah data yang digunakan dalam analisis sebaiknya memiliki ukuran yang lebih besar. Hal tersebut bertujuan untuk meningkatkan ketepatan klasifikasi. Bagi pihak RSUD Sidoarjo untuk memudahkan proses klasifikasi bayi baru lahir berisiko infeksi atau tidak diperlukan beberapa variabel lain, seperti kondisi ibu sebelum dan setelah melahirkan, kondisi lingkungan sekitar rumah yang tidak didapatkan dalam penelitian ini. Sehingga hasil klasifikasi bayi baru lahir terhadap status risiko infeksi akan menjadi lebih akurat.

## DAFTAR PUSTAKA

- [1] Departemen Kesehatan Republik Indonesia, *Laporan Nasional Riset Kesehatan Dasar Tahun 2007*. Jakarta: Depkes RI (2008).
- [2] Badan Pusat Statistik Provinsi Jawa Timur, *Laporan Eksekutif Kesehatan Provinsi Jawa Timur Tahun 2014*. Surabaya: BPS Jatim (2014).
- [3] L. Breiman, J. H Friedman, R. A. Olshen, dan C. J. Stone, *Classification and Regression Trees*. New York: Chapman Hall (1993).
- [4] R. J. Lewis, "An Introduction to Classification and Regression Trees (CART) Analysis," presented at the annual meeting of the Society for Academic Emergency Medicine, California. (2000).
- [5] I. H. Witten, E. Frank, dan M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed. United States: Morgan Kaufmann. (2011).
- [6] Kementerian Kesehatan RI, *Buku Saku Pelayanan Kesehatan Neonatal Esensial*. Jakarta: Kemenkes (2010).
- [7] WHO. *National Neonatal-Perinatal Database Report*. (2002).
- [8] WHO. *International Classification of Disease*. (2010).