

# MODEL POHON KEPUTUSAN PATOLOGI KLINIS PADA DIAGNOSIS PENYAKIT

Sri Mulyati

Jurusan Teknik Informatika, Fakultas Teknologi Industri  
Universitas Islam Indonesia, Yogyakarta

[sri.mulyati@uii.ac.id](mailto:sri.mulyati@uii.ac.id)

## Abstrak

*Pemeriksaan laboratorium setiap pasien di catat dalam data rekam medis. Seiring bertambahnya jumlah pasien maka data rekam medis juga semakin banyak. Banyaknya data rekam medis yang ada di rumah sakit akan menjadi lebih bermanfaat apabila dapat disarikan untuk menjadi pengetahuan. Penelitian ini tentang pembantuan pohon keputusan untuk pembentukan pohon keputusan pemeriksaan patologi klinis. Data yang digunakan adalah data pasien dengan diawali gejala demam dan didapatkan 10 penyakit. Pada dasarnya, proses penegakan diagnosis dilakukan melalui urutan yang jelas, yaitu dimulai dengan anamnesis, kemudian dilanjutkan dengan pemeriksaan fisik, dan pemeriksaan penunjang apabila diperlukan. Peran pemeriksaan laboratorium sangat diperlukan untuk mengkonfirmasi dan memastikan kondisi klinis pasien. Data rekam medis berupa pemeriksaan klinis diolah dengan algoritma ID3. Hasil dari penelitian ini berupa model pohon keputusan uji laboratorium dan penyakit.*

**Kata Kunci:** pohon keputusan, diagnosis, demam, algoritma ID3.

## 1. Pendahuluan

Proses penegakan diagnosis dilakukan melalui urutan yang jelas, yaitu dimulai dengan anamnesis, kemudian dilanjutkan dengan pemeriksaan fisik, dan pemeriksaan penunjang apabila diperlukan. Pemeriksaan penunjang meliputi pemeriksaan laboratorium dan radiologi. Peran pemeriksaan laboratorium dalam membuat keputusan klinis di antaranya dalam menegakkan diagnosis, monitor terapi, dan menentukan prognosis penyakit. Dalam menegakkan diagnosis, tidak cukup hanya mempertimbangkan anamnesis dan pemeriksaan fisik saja. Peran pemeriksaan laboratorium sangat diperlukan untuk mengkonfirmasi dan memastikan kondisi klinis pasien. Pemeriksaan laboratorium yang tepat dan analisis yang akurat sebenarnya sudah dapat digunakan sebagai penentu penyakit dalam proses diagnosis, di samping juga karena alasan ekonomis dan kemudahan (Speicher & Smith, 1994). Lee H. Hilborne dalam ASCP (2010) mengatakan bahwa "Diagnosis laboratorium hanya membutuhkan 1-2% dari seluruh biaya perawatan kesehatan, namun layanan laboratorium ini memberikan sumbangan paling banyak dalam memberikan dukungan keputusan".

Pemilihan uji laboratorium yang tidak tepat dan tidak lengkap dapat disebabkan oleh kesalahpahaman tentang hubungan suatu gejala atau tanda

dengan dengan data uji laboratorium. Karena dokter mengarahkan perhatian mereka pada data dalam konteks yang mereka anggap penting, berbagai data lain yang khas menjadi terabaikan (Wheeler, et al., 1977). Untuk merancang, menggunakan atau mengevaluasi secara efektif maka ciri-ciri diagnostik dari pemeriksaan laboratorium harus dipahami (Speicher & Smith, 1994).

Pemeriksaan laboratorium setiap pasien dicatat dalam data rekam medis. Seiring bertambahnya jumlah pasien maka data rekam medis juga semakin banyak. Banyaknya data rekam medis yang ada di rumah sakit akan menjadi lebih bermanfaat apabila dapat disarikan untuk menjadi pengetahuan.

## 2. Metode

Pembentukan pohon keputusan digunakan sebagai mesin inferensi untuk menganalisa fakta-fakta yang dimasukkan oleh pengguna. Metode ini dikenal dengan *backward chaining*, proses pencocokan fakta atau pernyataan dimulai dari bagian sebelah kanan (THEN dulu). Penalaran dimulai dari hipotesis terlebih dahulu, dan untuk menguji kebenaran hipotesis tersebut harus dicari fakta-fakta yang ada dalam basis pengetahuan. Data rekam medis pasien yang berupa uji laboratorium dan diagnosis diolah dengan metode ID3.

Pohon keputusan merupakan representasi dukungan keputusan yang diberikan secara grafis. Permasalahan yang didekati dengan menggunakan pohon keputusan bersifat saling bebas antara satu kejadian dengan kejadian yang lainnya. Pohon keputusan memiliki empat komponen utama, yaitu: akar (*root*), *node*, daun (*leaf*), dan busur (*arc*). Akar merupakan kejadian awal saat proses penelusuran akan dimulai. *Node* menunjukkan suatu atribut tertentu yang akan diuji kebenarannya. Daun menunjukkan klasifikasi, yaitu hasil yang diberikan setelah terjadi proses penelusuran mulai dari *root* hingga mencapai daun tersebut. Busur dengan anak panah menunjukkan arah penelusuran dari suatu kejadian ke kejadian berikutnya (Kusumadewi, et al., 2009).

Pemilihan suatu *node* menjadi *root*, dan menetapkan *node* mana yang akan menjadi alternatif berikutnya akan sangat menentukan efisiensi proses penelusuran. Untuk mengatasi permasalahan tersebut, dibutuhkan suatu mekanisme tertentu. Salah satu cara yang dapat dilakukan adalah dengan menguji dedikasi setiap atribut. Suatu atribut dianjurkan untuk menjadi *root* pada pohon keputusan, apabila atribut tersebut memiliki *information gain* paling tinggi.

Pemanfaatan *information gain* ini sangat terkait dengan konsep pembelajaran pada aturan yang disebut dengan induksi aturan (*rule induction*).

Misalkan atribut  $X$  memiliki  $m$  nilai, yaitu  $V_1, V_2, \dots, V_m$ . Probabilitas dari atribut bernilai  $V_j$  adalah:  $p(X = V_1) = p_1; p(X = V_2) = p_2; \dots; p(X = V_m) = p_m$ . Entropy dari  $X$ , selanjutnya akan disebut dengan  $H(X)$ , diberikan pada Persamaan 1. *Specific conditional entropy*, ditulis sebagai  $H(Y|X = v)$ , adalah *entropy* untuk memprediksi *output*  $Y$  apabila diberikan  $X = v$ . Sedangkan *conditional entropy*, ditulis sebagai  $H(Y|X)$  adalah rata-rata *specific conditional entropy* untuk  $Y$  (Persamaan 2). *Information gain*, selanjutnya ditulis sebagai  $IG(Y|X)$ , dihitung menggunakan Persamaan 3.

$$H(X) = -\sum_{j=1}^m p_j \log p_j = -p_1 \log p_1 - p_2 \log p_2 - \dots - p_m \log p_m \dots\dots\dots(1)$$

$$H(Y|X) = \sum_{j=1}^m p(X = v_j) H(Y|X = v_j) \dots\dots\dots(2)$$

$$IG(Y|X) = H(Y) - H(Y|X) \dots\dots\dots(3)$$

### 3. Pembahasan

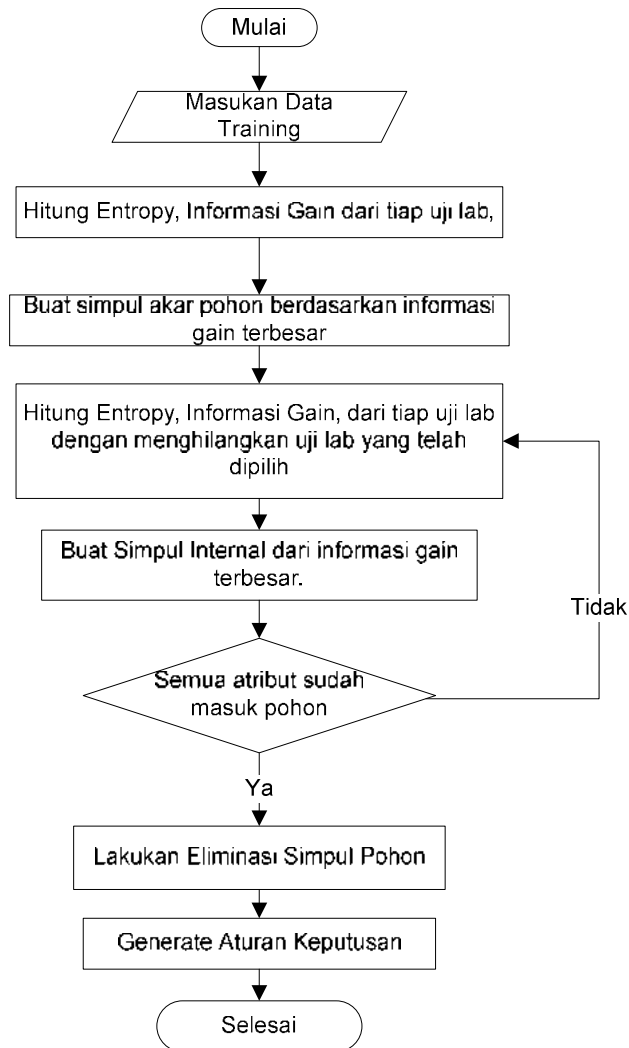
Proses perhitungan pohon keputusan digambarkan dalam *flowchart*. *Flowchart* merupakan bagan yang menunjukkan alur kerja atau apa yang sedang dikerjakan di dalam sistem secara keseluruhan dan menjelaskan urutan dari prosedur-prosedur yang ada di dalam sistem. Gambar 1 memperlihatkan *flowchart* Pembentukan Pohon Keputusan ID3. Berikut keterangan dari tiap-tiap proses pada *flowchart* pembentukan pohon keputusan dengan:

1. Data *training* dimasukan;
2. Menghitung *Information Gain* dan *Entropy* dari masing-masing uji laboratorium yang ada;
3. Membuat simpul akar dari uji laboratorium yang memiliki *Information Gain* terbesar;
4. Menghitung *Information Gain* dan *Entropy* dari masing-masing uji laboratorium dengan menghilangkan uji laboratorium yang telah dipilih sebelumnya;
5. Membuat simpul internal dari pemilihan uji laboratorium yang memiliki *Information Gain* terbesar;
6. Memeriksa apakah semua uji laboratorium sudah dibentuk pada pohon. Jika belum, maka ulangi proses d dan e, sedangkan jika sudah maka lanjut pada proses berikutnya;

7. Melakukan eliminasi simpul pohon untuk menghilangkan cabang-cabang yang tidak perlu; serta
8. Membuat aturan keputusan *degenerate* mengikuti aturan pohon yang telah terbentuk sebelumnya.

Catatan:

- $Entropy(S) = 0$ , Jika semua contoh pada  $S$  berada pada kelas yang sama;
- $Entropy(S) = 1$ , Jika semua contoh nilai atribut sama; dan
- $0 < Entropy(S) < 1$ , Jika jumlah contoh jika nilai atribut tidak sama.



**Gambar 1** Flowchart Pembentukan Pohon Keputusan

Pembentukan pohon keputusan diperoleh dengan melakukan perhitungan *Entropy* dan *Information Gain* pada data sampel uji laboratorium klinis. Data sampel uji laboratorium klinis tersebut diambil dari 10 penyakit yang dapat dilihat pada Tabel 1.

Tabel 1 Tabel Penyakit dan Jumlah Data

No	Kode ICD 10	Nama Penyakit	Jumlah
1	A01	Tyfoid Fever	10
2	A09	Diare	7
3	A90	Dengue Fever	10
4	B06.9	Rubella	7
5	J02	Faringitis Akut	9
6	J12	Pneumonia	16
7	J20	Bronkhitis	8
8	J21	Bronkhiolitis	6
9	J45.9	Asthma	4
10	P36	Sepsis	8
<b>Jumlah Data</b>			<b>85</b>

Sebagai contoh akan dilakukan perhitungan terhadap item uji laboratorium Hemoglobin (*Hb*) dan Hemakrit (*AT*) terhadap nilai item uji Normal, Kurang, Lebih dan *Not Available*. Berikut proses perhitungan *information gain*. Probabilitas penyakit dapat dihitung menggunakan Persamaan 4. *Information Gain* untuk Hemakrit dengan nilai Normal, Kurang, Lebih dan *Not Available* dapat dihitung dengan Persamaan 5 hingga 8.

$$\begin{aligned}
 H(Y) &= -\frac{10}{85} \log_2 \left( \frac{10}{85} \right) - \frac{7}{85} \log_2 \left( \frac{7}{85} \right) - \frac{10}{85} \log_2 \left( \frac{10}{85} \right) - \frac{7}{85} \log_2 \left( \frac{7}{85} \right) \\
 &\quad - \frac{9}{85} \log_2 \left( \frac{9}{85} \right) - \frac{16}{85} \log_2 \left( \frac{16}{85} \right) - \frac{8}{85} \log_2 \left( \frac{8}{85} \right) - \frac{6}{85} \log_2 \left( \frac{6}{85} \right) \\
 &\quad - \frac{4}{85} \log_2 \left( \frac{4}{85} \right) - \frac{8}{85} \log_2 \left( \frac{8}{85} \right) \dots \dots \dots (4) \\
 &= 3,2355
 \end{aligned}$$

$$\begin{aligned}
 H(Y | X_2 = \text{Kurang}) &= -\frac{2}{32} \log_2 \left( \frac{2}{32} \right) - \frac{1}{32} \log_2 \left( \frac{1}{32} \right) - \frac{1}{32} \log_2 \left( \frac{1}{32} \right) \\
 &\quad - \frac{4}{32} \log_2 \left( \frac{4}{32} \right) - \frac{11}{32} \log_2 \left( \frac{11}{32} \right) - \frac{3}{32} \log_2 \left( \frac{3}{32} \right) \\
 &\quad - \frac{4}{32} \log_2 \left( \frac{4}{32} \right) - \frac{6}{32} \log_2 \left( \frac{6}{32} \right) \dots \dots \dots (5) \\
 &= 2,6150
 \end{aligned}$$

$$\begin{aligned}
 H(Y | X_2 = \text{Lebih}) &= -\frac{2}{11} \log_2 \left( \frac{2}{11} \right) - \frac{7}{11} \log_2 \left( \frac{7}{11} \right) \\
 &\quad - \frac{1}{11} \log_2 \left( \frac{1}{11} \right) - \frac{1}{11} \log_2 \left( \frac{1}{11} \right) \dots \dots \dots (6) \\
 &= 1,4911
 \end{aligned}$$

$$\begin{aligned}
 H(Y | X_2 = \text{Normal}) &= -\frac{1}{28} \log_2 \left( \frac{1}{28} \right) - \frac{2}{28} \log_2 \left( \frac{2}{28} \right) - \frac{5}{28} \log_2 \left( \frac{5}{28} \right) \\
 &\quad - \frac{3}{28} \log_2 \left( \frac{3}{28} \right) - \frac{3}{28} \log_2 \left( \frac{3}{28} \right) - \frac{2}{28} \log_2 \left( \frac{2}{28} \right) \\
 &\quad - \frac{5}{28} \log_2 \left( \frac{5}{28} \right) - \frac{3}{28} \log_2 \left( \frac{3}{28} \right) - \frac{4}{28} \log_2 \left( \frac{4}{28} \right) \dots \dots \dots (7) \\
 &= 3,04007
 \end{aligned}$$

$$\begin{aligned}
 H(Y | X_2 = \text{N / A}) &= -\frac{1}{14} \log_2 \left( \frac{1}{14} \right) - \frac{2}{14} \log_2 \left( \frac{2}{14} \right) - \frac{2}{14} \log_2 \left( \frac{2}{14} \right) \\
 &\quad - \frac{3}{14} \log_2 \left( \frac{3}{14} \right) - \frac{2}{14} \log_2 \left( \frac{2}{14} \right) - \frac{3}{14} \log_2 \left( \frac{3}{14} \right) \\
 &\quad - \frac{1}{14} \log_2 \left( \frac{1}{14} \right) \dots \dots \dots (8) \\
 &= 2,69951
 \end{aligned}$$

$$H(Y | X_2) = \frac{32}{85} \cdot 2,6150 + \frac{11}{85} \cdot 1,4911 + \frac{28}{85} \cdot 3,04007 + \frac{14}{85} \cdot 2,69951 \dots\dots\dots(9)$$

$$= 2,62$$

$$IG(Y | X_2) = 3,236 - 2,62 = 0,612 \dots\dots\dots(10)$$

Dari data sampel yang ada didapatkan hasil perhitungan *information gain*-nya, seperti pada Tabel 2.

**Tabel 2** Tabel Hasil Perhitungan *Information Gain*

No	Variabel	<i>Information Gain</i>	No	Variabel	<i>Information Gain</i>
1	Limfosit	0.6467	18	Silinder	0.2376
2	Leukosit	0.6259	19	Kristal Urine	0.2376
3	Hemaktorit	0.612	20	BJ Urine	0.2326
4	Trombosit	0.5682	21	MCH	0.1602
5	Monosit	0.4796	22	LED 2 Jam	0.1537
6	Eritrosit(AE)	0.4636	23	Keton	0.1317
7	Potein Urin	0.4314	24	MCHC	0.1061
8	Kekeruhan Urine	0.4119	25	Netrofil Barang	0.0986
9	UrineLeukosit	0.371	26	Glukosa Sewaktu	0.0969
10	Hemoglobin	0.3687	27	Natrium	0.0609
11	LED 1 Jam	0.3351	28	Klorida	0.0609
12	PH Urine	0.3162	29	Kalium	0.060
13	Warna	0.3104	30	Limfosit B	0.0545
14	Epitel Urine	0.3104	31	Bilirubin	0.039
15	Bakteri	0.2744	32	Urobilinogen	0.0371
16	MCV	0.2707	33	Ureum	0.0371
17	Eritrosit (AE)	0.266	34	Kreatinin	0.0371

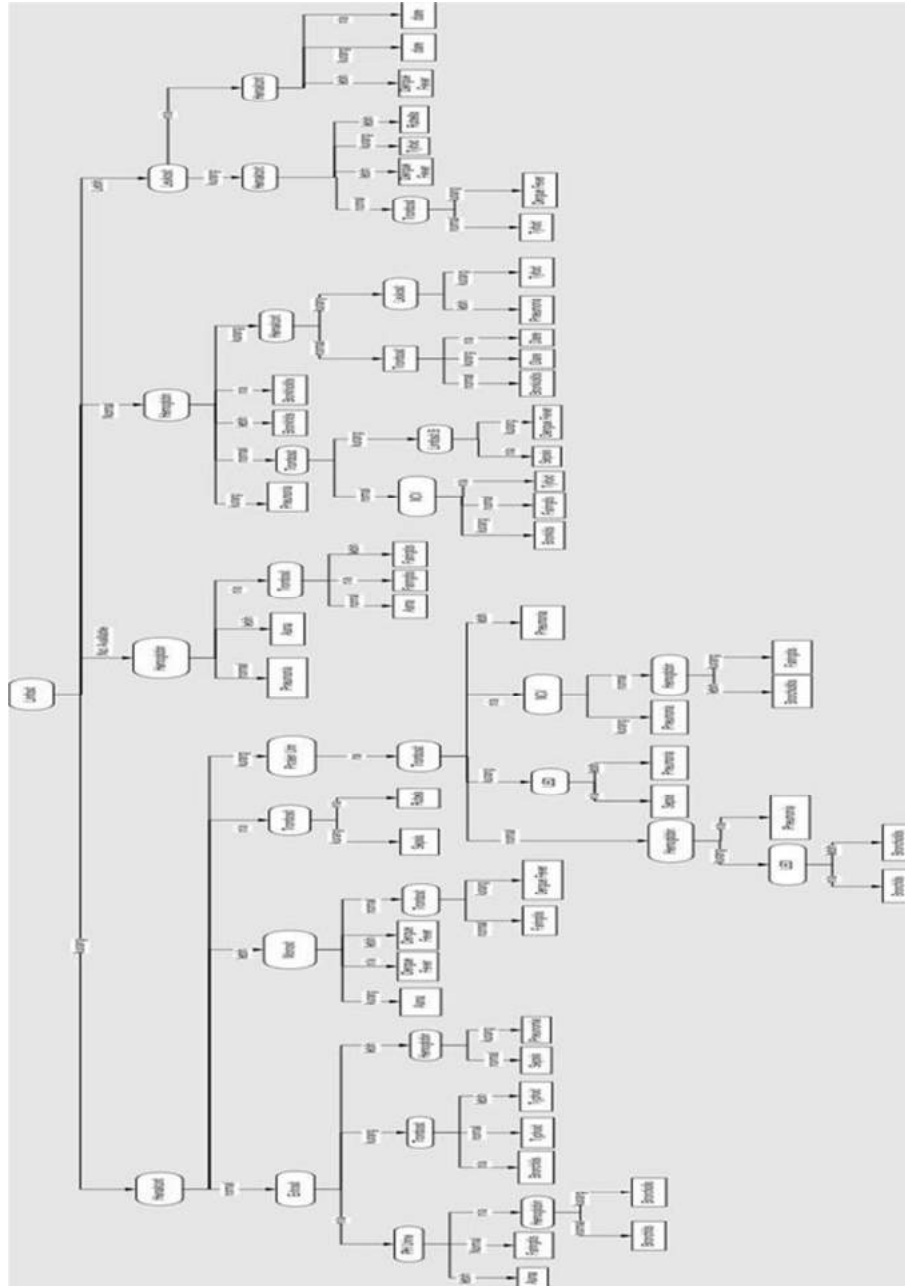
Penelitian ini menggunakan 34 jenis pemeriksaan. Dari pemeriksaan tersebut dihasilkan *information gain* tertinggi, yaitu Limfosit *T*. Oleh karena itu maka Limfosit *T* dijadikan sebagai *root*. Gambar 2 memperlihatkan Diagram Pembentukan Pohon Keputusan.

#### 4. Hasil

Setelah dilakukan pengujian sistem, terdapat beberapa kelebihan dan kelemahan pada sistem. Kelebihan-kelebihan yang ada pada sistem, antara lain yaitu: Sistem dapat membentuk pohon keputusan berdasarkan basis pengetahuan dari uji laboratorium yang ada.

Sedangkan kelemahan yang masih terdapat pada sistem yaitu:

1. Kurangnya basis pengetahuan yang dimiliki sehingga masih terdapat hasil pengolahan model keputusan kedua kurang maksimal dalam pembentukan pohon keputusan.
2. Kurangnya data kasus rekam medis menjadikan penyebaran penyakit tidak seimbang.



Gambar 2 Diagram pembentukan pohon keputusan

**5. Penutup**

Model pohon keputusan dapat diterapkan sebagai basis pengetahuan untuk menentukan jenis uji laboratorium yang relevan berdasarkan diagnosis awal.

1. Perlu adanya penambahan data penyakit selain Typhoid Fever, Diare, Dengue Fever, Rubela, Faringitis Akut, Pneumonia, Bronkhitis, Bronkiolitis, Asma dan Sepsis.

2. Perlu dilakukan *preprocessing* lebih ketat untuk menyaring data uji laboratorium pada saat pengumpulan data.
3. Perlu penambahan data kasus untuk pengujian agar nilai kinerja sistem dapat meningkat.
4. Partisipasi dokter perlu diintensifkan dalam rangka melakukan uji coba atas kinerja sistem ini.

### Daftar Pustaka

- ASCP, 2010. *Pathology's Future: A View from Leaders in Health Care*. [Online] Available at: <http://www.ascpresources.org/e-books/future/files/future.pdf> [Accessed 17/08/2011].
- Bemmel, J. H. & Musen, M. A., 1997. *Modelling of Decision Support in Handbook of Medical Informatics*. Diegem, Belgium: Bohn Stafleu Van Loghum.
- Kusumadewi, S., Fauziah, A., Khoiruddin, A. A., Wahid, F., Setiawan, M. A., Rahayu, N. W., Hidayat, T. & Prayudi, Y., 2009. *Informatika Kesehatan*. Yogyakarta: Graha Ilmu.
- Mulyati, S. & Kusumadewi, S., 2012. Model Sistem Pendukung Keputusan Untuk Diagnosis Penyakit Anak Dengan Gejala Demam Menggunakan Naive Bayesian Classification. *Proceeding at Seminar Nasional Informatika Medis (SNIMed) III 2012*.
- Speicher, C. E. & Smith, J. W., 1994. *Pemilihan Uji Laboratorium yang Efektif: Choosing Effective Laboratory Tests*, Edisi Terjemahan. Jakarta: Penerbit Buku Kedokteran EGC.
- Wheeler, L. A., Brecher, G. & Sheiner, L. B., 1977. Clinical Laboratory Use in the Evaluation of Anemia. *Journal of the American Medical Association (JAMA)*, 238(25), pp. 2709-2714.