

## DIAGNOSA TINGKAT KESEHATAN PASIEN MENGUNAKAN METODE DECISION TREE

Hidayatulah Himawan<sup>1</sup>, Oliver S. Simanjuntak<sup>2</sup>, Agus Triawan<sup>3</sup>

<sup>1,2,3</sup>Program Studi Teknik Informatika

UPN "Veteran" Yogyakarta

Jl. Babarsari No 2, Tambakbayan, Yogyakarta

e-mail : if.iwan@gmail.com<sup>1</sup>; oliversimanjuntak@yahoo.com<sup>2</sup>

### Abstrak

*Keadaan kesehatan merupakan keinginan tiap orang untuk selalu memiliki tingkat dan kualitas kesehatan yang baik. Munculnya berbagai gejala penyakit yang menyebabkan seseorang menderita suatu penyakit dapat diketahui secara dini. Untuk itu perlu adanya suatu aplikasi pendukung secara online baik berupa web online ataupun aplikasi lainnya yang bisa digunakan secara cepat dan tepat untuk diketahui, sebelum penyakit tersebut ditangani oleh orang yang ahli dibidangnya (dokter). Aplikasi ini untuk mendiagnosa tingkat kesehatan seseorang khususnya yang aktif sebagai perokok. Aplikasi ini menggunakan metode decision tree dengan kombinasi algoritma C.45 yang menggambarkan tingkat kualitas kesehatan seseorang akibat bahaya merokok yang ditimbulkan. Sampling diambil pada 70 orang pasien dengan metode wawancara dan kuesioner yang memberikan efek dan dampak serta kebiasaan dari pasien tersebut.*

**Kata Kunci :** Web Online, Decision Tree, Aplikasi.

### 1. PENDAHULUAN

Pada umumnya semua orang pasti menginginkan kualitas kesehatan secara menyeluruh terhadap kehidupan mereka secara nyata. Namun bukan rahasia lagi jika menjaga kesehatan lebih sulit ketika orang tersebut belum merasakan dampak dan akibat dari pola hidup yang kurang baik dan kurang teratur. Salah satu yang bisa dilihat adalah kebiasaan seseorang untuk merokok. Mereka telah mengetahui bahaya rokok secara langsung, hal ini bias terjadi dikarenakan dalam setiap kemasan rokok telah tertera tulisan peringatan bahaya rokok untuk kesehatan (Kusyogo, dkk, 2012), Didalam rokok terdapat lebih dari 4000 zat dan 2000 diantaranya zat tersebut merupakan zat yang berdampak bagi kesehatan. Diantara zat-zat berbahaya tersebut bahan radioaktif (*polonium-201*). dan bahan yang dipakai dalam cat (*acetone*). Obat gegat (*naphthalene*), ada juga racun untuk serangga, pencuci lantai juga terdapat pada rokok atau sering disebut juga *ammonia*, racun *arsenic* atau racun untuk anai-anai, gas beracun (*hydrogen cyanide*). Dengan semakin tingginya tingkat kesadaran manusia atas kesehatan bagi diri mereka dan lingkungan disekitar mereka, maka pada saat ini mulai bermunculan kampanye-kampanye aktif untuk mengurangi serta mengingatkan para pengguna rokok akan bahaya yang dapat ditimbulkannya kelak (Akbar AA, 2012).

Banyak faktor yang mempengaruhi para pengguna bagi resiko bahaya merokok khususnya bagi kesehatan jantung (Mamat, 2008). Beberapa faktor yang dapat mempengaruhi adalah lingkungan dan kebiasaan seseorang untuk merokok. Perlu adanya perhatian serius tidak hanya oleh keluarga namun pembatasan dan penerapan aturan bagi perokok harus mulai ditingkatkan. Pengelompokan terhadap para perokok perlu memperhatikan klasifikasi dan pola yang ada (Edy K, dkk, 2009). Tujuan pengelompokan ini lebih kepada cara di masa yang akan datang untuk menangani dan menanggulangi tiap pasien yang ada.

Metode decision tree dengan algoritma C.45 dapat digunakan untuk memprediksi berbagai informasi dari data yang akan digunakan (Angga, 2009). Pada aplikasi ini, diagnose yang dikembangkan akan lebih mengarah kepada penyakit tertentu, khususnya bronchitis, sehingga pasien yang didapatkan akan lebih terdeteksi secara dini.

### 2. TINJAUAN PUSTAKA

Wenny W, dkk (2012), melakukan riset terhadap penderita tuberculosis dengan mengembangkan aplikasi sistem pakar untuk mendeteksi secara dini penyakit TBC pada diri pasien. Metode waterfall digunakan untuk melihat data awal yang telah diperoleh dari para tenaga ahli (dokter). Aplikasi dikembangkan dengan PHP dan MySQL. Ellyza G (2013) mengembangkan sebuah prototype sistem pakar untuk mendeteksi tingkat resiko penyakit jantung koroner dengan metode Dempster-Shafer. Penelitian lainnya dilakukan oleh Rachmawati dkk (2012) dengan mengembangkan aplikasi sistem pakar diagnosa penyakit asma.

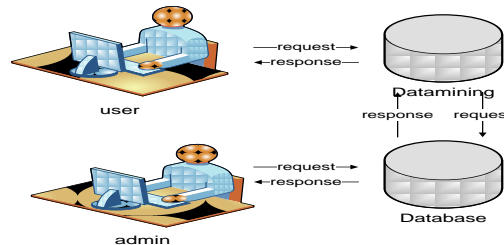
### 3. METODE PENELITIAN

Metode waterfall digunakan untuk melaksanakan penelitian ini. Data diambil melalui metode sampling yang melibatkan 70 responden dengan cara wawancara. Data diolah menggunakan Decision Tree untuk mendapatkan data yang lebih akurat. Pada tahap analisis melihat kebutuhan sistem yang ada. Mulai dari sisi kebutuhan sistem yang dapat mendukung penggunaan aplikasi, hingga kebutuhan user terhadap aplikasi. Dari analisa tersebut akan dikembangkan sebuah rancangan yang sesuai dengan kebutuhan pengguna. Rancangan yang ada akan diimplementasikan ke dalam sebuah aplikasi untuk mendiagnosa tingkat kesehatan seseorang terutama terhadap bahaya rokok. Algoritma C.45 digunakan untuk melihat perbandingan data dalam implementasi untuk memprediksi pola dan data yang akan terbentuk.

### 4. HASIL DAN PEMBAHASAN

#### Identifikasi Sistem

Identifikasi sistem merupakan gambaran model diagram untuk menunjukkan tata letak sebuah sistem secara fisik dengan menampilkan bagian-bagian perangkat lunak (*software*) yang berjalan pada perangkat keras (*hardware*). Pada aplikasi ini perangkat keras yang digunakan adalah *laptop* dengan sistem operasi *Windows7* untuk sarana penerapan aplikasi yang akan dibangun. Aplikasi ini akan dibangun dengan *Notepad++*. Aplikasi ini menjadikan *MySQL* sebagai *database*-nya. *Admin* menginputkan data yang kemudian disimpan dalam database. Ketika *user* menjalankan aplikasi, data akan diambil dari database yang kemudian akan diproses ke dalam data *mining*. Untuk menggambarkan arsitektur sistemnya bisa dilihat pada gambar dibawah.



Gambar 1. Arsitektur Sistem Aplikasi Data Mining

#### Langkah -1: Mengubah Data Menjadi Tree. Menghitung Data Pasien Rumah Sakit.

Tabel 1. Jumlah Data Penyakit Bronkhitis

Penyakit	Terkena Penyakit	Tidak Terkena Penyakit	Total
Penyakit Bronkhitis	37	33	70

#### Langkah penyelesaian kasus:

1. Mengubah data menjadi tree.
2. Mengubah tree menjadi rule.

Namun pada penelitian ini hanya sampai langkah 1 dengan mengubah data menjadi tree.

Menentukan Node Terpilih, Dari data sampel tentukan dulu node terpilih yaitu dengan menghitung nilai information gain masing-masing atribut. (usia, berat badan, dan jenis kelamin).

Catatan:

1. Untuk menentukan node terpilih, gunakan nilai *information gain* dari setiap kriteria dengan data sample yang ditentukan.
2. Node terpilih adalah kriteria dengan *information gain* yang paling besar.

Langkah – langkah menentukan node terpilih:

#### 1. Menghitung nilai informasi (I) dari seluruh data training

Dengan menggunakan persamaan:

$$-p(+)\log_2 p(+)-p(-)\log_2 p(-)$$

$$= -37/70\log_2 37/70 - 33/70\log_2 33/70 = 0,99$$

#### 2. Menghitung nilai informasi tiap atribut di ambil dari satu penyakit bronkhitis. Dengan menggunakan persamaan:

$$-p(+)\log_2 p(+)-p(-)\log_2 p(-)$$

**Tabel 1. Data Jumlah Rokok Per-Hari**

Jumlah Rokok	Bronkhitis	Jumlah
2-11 Batang	YA	10
2-11 Batang	TIDAK	15
12-24 Batang	YA	26
12-24 Batang	TIDAK	15
25-36 Batang	YA	1
25-36 Batang	TIDAK	3
	<b>Total</b>	70

Hitung Data Jumlah Rokok Per-Hari:

$$Q1 = -10/25 \log_2 10/25 - 15/25 \log_2 15/25 = 0,97$$

$$Q2 = -26/41 \log_2 26/41 - 15/41 \log_2 15/41 = 0,94$$

$$Q3 = -1/4 \log_2 1/4 - 3/4 \log_2 3/4 = -0,05$$

Hitung Jumlah Data Keluhan Merokok:

Sesak Napas:

$$Q1 = -26/50 \log_2 26/50 - 24/50 \log_2 24/50 = 0,99$$

$$Q2 = -11/20 \log_2 11/20 - 9/20 \log_2 9/20 = 0,99$$

Hitung Data Jumlah Batuk:

$$Q1 = -20/38 \log_2 20/38 - 18/38 \log_2 18/38 = 0,99$$

$$Q2 = -17/32 \log_2 17/32 - 15/32 \log_2 15/32 = 0,99$$

Hitung Data Jumlah Dahak:

$$Q1 = -11/17 \log_2 11/17 - 6/17 \log_2 6/17 = 0,93$$

$$Q2 = -26/53 \log_2 26/53 - 27/53 \log_2 27/53 = 0,99$$

Hitung Data Jumlah Nyeri Dada:

$$Q1 = -10/18 \log_2 10/18 - 8/18 \log_2 8/18 = 0,99$$

$$Q2 = -27/52 \log_2 27/52 - 25/52 \log_2 25/52 = 0,99$$

Hitung Data Jumlah Pusing:

$$Q1 = -6/11 \log_2 6/11 - 5/11 \log_2 5/11 = 0,99$$

$$Q2 = -31/59 \log_2 31/59 - 28/59 \log_2 28/59 = 0,99$$

Hitung Data Jumlah Flu:

$$Q1 = -7/12 \log_2 7/12 - 5/12 \log_2 5/12 = 0,97$$

$$Q2 = -30/58 \log_2 30/58 - 28/58 \log_2 28/58 = 0,99$$

Hitung Data Jumlah Panas:

$$Q1 = -8/14 \log_2 8/14 - 6/14 \log_2 6/14 = 0,98$$

$$Q2 = -29/56 \log_2 29/56 - 27/56 \log_2 27/56 = 0,01$$

Hitung Jumlah Data Lama rokok:

$$Q1 = -12/16 \log_2 12/16 - 4/16 \log_2 4/16 = 0,97$$

$$Q2 = -11/19 \log_2 11/19 - 8/19 \log_2 8/19 = 0,98$$

$$Q3 = -9/21 \log_2 9/21 - 12/21 \log_2 12/21 = 0,98$$

$$Q4 = -5/14 \log_2 5/14 - 9/14 \log_2 9/14 = 0,94$$

Hitung Jumlah Data Pengalaman Berhenti Merokok:

$$Q1 = -20/50 \log_2 20/50 - 29/50 \log_2 29/50 = 0,97$$

$$Q2 = -17/20 \log_2 17/20 - 4/20 \log_2 4/20 = 0,70$$

### 3. Menghitung nilai *entropy* tiap atribut:

**Jumlah Rokok Per-hari:**

$$E = (25/70) Q1 + (41/70) Q2 + (4/70) Q3 = -0,2$$

**Keluhan:**

a. Sesak Napas.

$$E = (50/70) Q1 + (20/70) Q2 = -0,27$$

b. Batuk

$$E = (38/70) Q1 + (32/70) Q2 = -0,27$$

c. Dahak

$$E = (17/70) Q1 + (53/70) Q2 = -0,25$$

d. Nyeri Dada

$$E = (18/70) Q1 + (52/70) Q2 = -0,27$$

e. Pusing

$$E = (11/70) Q1 + (59/70) Q2 = -0,27$$

f. Flu

$$E = (12/70) Q1 + (58/70) Q2 = -0,26$$

- g. Panas  
 $E = (14/70) Q1 + (56/70) Q2 = -0,52$
- h. Lama Rokok  
 $E = (16/70) Q1 + (19/70) Q2 + (21/70) Q3 + (14/70) Q4 = -0,24$
- i. Pengalaman Berhenti Merokok  
 $E = (49/70) Q1 + (21/70) Q2 = -0,16$

**Menghitung nilai *information gain* tiap atribut.**

**Tabel 2. Hasil Data Gain Penyakit Brokhitis**

Data	Nilai Seluruh Data	Nilai Entropy Tiap Atribut	Jumlah
Jumlah Rokok	0,99	0,92	0,07
Sesak Napas	0,99	0,99	0
Batuk	0,99	0,99	0
Dahak	0,99	0,97	0,02
Nyeri Dada	0,99	0,99	0
Pusing	0,99	0,99	0
Flu	0,99	0,98	0,01
Panas	0,99	0,2	0,79
Lama Rokok	0,99	0,96	0,03
Pengalaman Berhenti	0,99	0,9	0,09

Dengan menggunakan langkah – langkah yang sama di hitung nilai *information gain* atribut seluruhnya, sehingga didapat nilai *information gain* atribut Panas sebesar 0,79 Sehingga, terpilih atribut

## 5. KESIMPULAN

Dari hasil penelitian ini, maka diambil kesimpulan sebagai berikut :

1. Jumlah rokok perhari di dapat  $Q1 = 0,97$ ,  $Q2 = 0,94$ ,  $Q3 = -0,05$
2. Dengan Entropy atribut yang dihasilkan sebesar -0,2 maka proses penghitungan jumlah rokok perhari dapat dihasilkan dengan entropy tersebut.
3. Aplikasi untuk mendiagnosa tingkat kesehatan pasien menghasilkan *information gain* atribut panas sebesar 0,79. Sehingga proses analisa tingkat kesehatan diakibatkan oleh penyakit dahak yang lebih mengarah kepada bronkhitis.

## DAFTAR PUSTAKA

- Edy Kurniawan, I Ketut Edy Purnama, Surya Sumpeno, “*Analisa Rekam Medis untuk Menentukan Pola Kelompok Penyakit Menggunakan Klasifikasi dengan Decision Tree J4*”, Institut Sepuluh Nopember, Surabaya.
- Akbar Aji Anugraha, 2012, “*Sistem Pakar Deteksi Bahaya Rokok Bagi Kesehatan*” Universitas Dian Nuswantoro Semarang.
- Angga Raditya, 2009, *Implementasi Data Mining Classification untuk Mencari Pola Prediksi Hujan dengan Menggunakan Algoritma C4.5*, Universitas Gunabangsa, Depok.
- Ellyza Gustri Wahyuni dan Widodo Prijodiprojo, 2013 “*Prototype Sistem Pakar untuk Mendeteksi Tingkat Resiko Penyakit Jantung Koroner dengan Metode Dempster-Shafer*”, FMIPA UGM, Yogyakarta.
- Wenny Widiastuti, Dini Destiani, Dhani Johar Damiri, 2012, “*Aplikasi Sistem pakar deteksi dini pada penyakitTuberkulosis*”, Sekolah Tinggi Teknologi Garut.
- Rachmawati, Dhani Johar Damiri, Ate Susanto, 2012, “*Aplikasi Sistem Pakar Diagnosis Penyakit Asma*” Sekolah Tinggi Teknologi Garut.
- Kusyogo Cahyo, Putri Asmita Wigati, Zahroh Shaluhiah, 2012 “*Rokok, Pola Pemasaran dan Perilaku Merokok Siswa SMA/Sederajat di Kota Semarang*, Universitas Diponegoro Semarang.
- Mamat, 2008, “*Faktor-faktor resiko yang berpengaruh terhadap kejadian penyakit jantung koroner terhadap kelompok usia <45 Tahun*”. Unoversitas Diponegoro Semarang