

## **PENERAPAN DATA MINING UNTUK KLASIFIKASI PREDIKSI PENYAKIT ISPA (*Infeksi Saluran Pernapasan Akut*) DENGAN ALGORITMA DECISION TREE (ID3)**

<sup>1</sup>Aline Embun Pramadhani (08018360), <sup>2</sup>Tedy Setiadi (0407016801)

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

Universitas Ahmad Dahlan

Prof. Dr. Soepomo, S.H., Janturan, Umbulharjo, Yogyakarta 55164

<sup>1</sup>Email: aline.embun11@gmail.com

<sup>2</sup>Email: tedy.setiadi@tif.uad.ac.id

### **ABSTRAK**

Pengolahan data yang disimpan tidak hanya bisa disimpan saja, tetapi bisa dijadikan suatu representasi pengetahuan dikemudian hari. Data pasien yang terkena Ispa digunakan untuk merepresentasikan pengetahuan dari gejala penyakit ispa sebelumnya, yang mana diderita oleh pasien di klinik pengobatan Dharma Husada. Dengan demikian, perlu adanya suatu klasifikasi penyakit yang paling banyak diderita pasien di klinik Dharma Husada.

Klasifikasi pada penelitian ini bertujuan untuk membentuk suatu model pohon keputusan untuk memprediksi penyakit pasien dan melihat variable yang paling mempengaruhi penyakit pasien dengan kategori ispa. Obyek penelitian ini adalah hasil data pasien, status imunisasi, jenis kelamin, usia dan kriteria gizi pada tahun 2012 yaitu sebanyak 200 pasien terkena ispa. Data yang digunakan adalah data pemeriksaan pasien oleh dokter untuk kemudian dapat dinyatakan terkena penyakit ispa.

Variable bebas atau variable *input* (predictor) pada penelitian ini adalah pemeriksaan awal pasien yaitu meliputi kategori status imuniasasi, jenis kelamin, usia dan kriteria gizi. Sedangkan kondisi pasien yang diprediksi yaitu terkena ispa atau tidak merupakan *output* (variable tidak bebas). Tools yang digunakan untuk membuat aplikasi penerapan data mining adalah dengan VB.Net dengan pengolahan database SQL server. Pengujian terhadap pembentukan pohon keputusan menggunakan uji Chi Square untuk mengetahui nilai hipotesis hitung yang dibandingkan dengan tabel chi square. Dari pohon keputusan yang terbentuk dari 200 data pasien maka dapat diketahui bahwa jenis kelamin tidak berpengaruh terhadap penyakit ispa.

Kata kunci : *Pohon Keputusan, Decision Tree, Chi Square, Penyakit Ispa.*

### **1. PENDAHULUAN**

Teknologi informasi saat ini, sudah digunakan pada berbagai instansi yang memerlukan pengolahan data yang banyak. Misalnya, memasukan data,

mengolah data dan kemudian dijadikan sebagai suatu informasi yang bermanfaat. Data dalam jumlah banyak pada masa lalu, dapat digunakan sebagai informasi pada masa yang akan datang. Data tersebut dapat diolah dengan menggunakan suatu metode untuk menggali informasi yang sudah ada misalnya untuk mengetahui kategori gaji pegawai berdasarkan jabatan, umur dan jenis kelamin. Dengan adanya data tersebut, maka dapat dilakukan analisis untuk mengetahui masing-masing kategori gaji pegawai.

Balai Pengobatan dan Klinik Dharma Husada adalah sebuah tempat pengobatan baik rawat inap ataupun rawat jalan, yaitu sebuah instansi yang memiliki banyak data pasien, seperti nama, usia, jenis kelamin, dan gejala suatu penyakit yang disimpan dalam suatu database. Kesulitan yang dirasakan oleh pihak balai pengobatan adalah dalam mengklasifikasikan seorang pasien menderita penyakit ISPA atau tidak. ISPA adalah salah satu penyakit yang sering diderita pasien di sekitar balai pengobatan yaitu di daerah Parangtritis Bantul, dengan berbagai gejala ISPA yang kemudian tidak diketahui faktor apa saja yang memiliki kecenderungan balita terkena ISPA, sebesar 60% balita di daerah parangtritis terkena ISPA. Karena gejala yang dialami berbeda-beda dari pasien satu dengan yang lain, dan kemudian hanya didiagnosa dengan penyakit ISPA.

Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui. Dengan bantuan teknik data mining dapat diketahui pola suatu penyakit berdasarkan data yang sudah ada, yaitu nama pasien, usia, jenis kelamin, status imunisasi dan status gizi dengan gejala penyakit ISPA (pernapasan). Dengan demikian, jika sudah diketahui faktor-faktor yang mempengaruhi suatu diagnose, maka memudahkan untuk klasifikasi untuk pola keputusan suatu penyakit. [4]

Metode yang digunakan dalam klasifikasi keputusan tersebut adalah metode decision tree (ID3), yaitu pohon keputusan yang memperlihatkan faktor-faktor kemungkinan atau probabilitas yang akan mempengaruhi alternatif-alternatif keputusan tersebut, disertai dengan estimasi hasil akhir yang akan didapat bila kita mengambil alternatif keputusan tersebut.

## 2. KAJIAN PUSTAKA

Metode Iterative Dichtomizer 3 (ID3) atau lebih sering dikenal dengan pohon keputusan adalah penelitian yang memiliki tujuan untuk penyeleksian penerimaan mahasiswa baru, dalam pengambilan keputusan penerimaan mahasiswa baru ini dibatasi dengan memperhatikan tiga atribut yaitu nilai SPMB, UAN, dan psikotest. Metode yang digunakan adalah metode *Decision Tree* menggunakan ID3 (*Iterative Dichotomiser 3*). Pengambilan keputusan ini dapat memberikan data yang lengkap sehingga akan lebih mudah bagi pihak universitas menentukan seseorang atau peserta SPMB (Seleksi Penerimaan Mahasiswa Baru) diterima atau tidak di universitas tersebut. [1]

Penelitian yang dilakukan oleh Anita Prawitasari (2011) yaitu mengklasifikasikan yang bertujuan untuk membentuk suatu model pohon keputusan untuk memprediksi penyakit pasien dengan kategori tipes dan DBD. Obyek penelitian ini adalah hasil uji laboratorium pasien, ciri-ciri, gejala dan kondisi fisik pasien yang didapat dan rekam medis pusat RSCM pada tahun 2009 yaitu sebanyak 47 data pasien, 20 pasien yang tipes dan 27 pasien yang DBD. Data yang digunakan adalah hasil laboratorium pasien yang diambil pada saat pemeriksaan pertama pasien datang sebelum dinyatakan memiliki penyakit tipes atau DBD. Variable bebas atau variable *input* (predictor) pada penelitian ini adalah pemeriksaan awal pasien yaitu meliputi kategori usia, demam, lama demam, waktu demam, keadaan fisik, ruam atau bintik, warna lidah, trombosit, hematokrit dan uji widal pasien. Sedangkan konsisi pasien yang diprediksi yaitu status pasien tipes atau DBD merupakan output (variable tidak bebas). [2]

### 2.1 Data Mining

Secara sederhana, *data mining* atau penambangan data dapat didefinisikan sebagai proses seleksi, eksplorasi, dan pemodelan dari sejumlah besar data untuk menemukan pola atau kecenderungan yang biasanya tidak disadari keberadaannya. *Data mining* dapat dikatakan sebagai proses mengekstrak pengetahuan dari sejumlah besar data yang tersedia. Pengetahuan yang dihasilkan dari proses *data mining* harus baru, mudah dimengerti, dan bermanfaat. Dalam *data mining*, data disimpan secara elektronik dan diproses secara otomatis oleh komputer menggunakan teknik dan perhitungan tertentu.[5]

### 2.1 Algoritma Decision Tree

contoh tabel konsep data [3]

Nama	Usia	Berat	Kelamin	Hipertensi
Ali	Muda	Overweight	Pria	Ya
Edi	Muda	Underweight	Pria	Tidak
Annie	Muda	Average	Wanita	Tidak
Budiman	Tua	Overweight	Pria	Tidak
Herman	Tua	Overweight	Pria	Ya
Didi	Muda	Underweight	Pria	Tidak
Nina	Tua	Overweight	Wanita	Ya
Gatot	Tua	Average	Pria	Tidak

memilih node awal

Usia	Hipertensi	Jumlah
Muda	Ya (+)	1
Muda	Tidak (-)	3
Tua	Ya	2
Tua	Tidak	2

Keterangan :  
Usia = muda

$$q_1 = -\frac{1}{4}\log_2\frac{1}{4} - \frac{3}{4}\log_2\frac{3}{4} = 0.81$$

Usia = tua

$$q_1 = -\frac{2}{4}\log_2\frac{2}{4} - \frac{2}{4}\log_2\frac{2}{4} = 1$$

Entropi untuk usia =

$$E = \frac{4}{8}q_1 + \frac{4}{8}q_2 = \frac{4}{8}(0.81) + \frac{4}{8}(1) = 0.91$$

Melakukan perhitungan sesuai dengan kriteria

Usia	Hipertensi	Jumlah	Kelamin	Hipertensi	Jumlah
Muda	Ya	1	Pria	Ya	2
Muda	Tidak	3	Pria	Tidak	4
Tua	Ya	2	Wanita	Ya	1
Tua	Tidak	2	Wanita	Tidak	1

Penyelesaian :

Usia = Muda

$$q_1 = -\frac{1}{4}\log_2\frac{1}{4} - \frac{3}{4}\log_2\frac{3}{4} = 0.81$$

Entropi untuk usia =

$$E = \frac{4}{8}q_1 + \frac{4}{8}q_2$$

Usia = Tua

$$= \frac{4}{8}(0.81) + \frac{4}{8}(1) = 0.91$$

$$q_2 = -\frac{2}{4}\log_2\frac{2}{4} - \frac{2}{4}\log_2\frac{2}{4} = 1$$

Jenis Kelamin = Pria

$$q_1 = -\frac{2}{6}\log_2\frac{2}{6} - \frac{4}{6}\log_2\frac{4}{6} = 0.91$$

Entropi untuk usia =

$$E = \frac{6}{8}q_1 + \frac{2}{8}q_2$$

Jenis Kelamin = Wanita

$$= \frac{6}{8}(0.91) + \frac{2}{8}(1) = 0.93$$

$$q_2 = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1$$

Tabel berdasarkan kriteria berat badan

Berat	Hipertensi	Jumlah
Overweight	Ya	3
Overweight	Tidak	1
Average	Ya	0
Average	Tidak	2
Underweight	Ya	0
Underweight	Tidak	2

Penyelesaian

Berat badan = Overweight

$$q_1 = -\frac{3}{4}\log_2\frac{3}{4} - \frac{1}{4}\log_2\frac{1}{4} = 0.81$$

Entropi untuk berat badan

$$E = \frac{4}{8}q_1 + \frac{2}{8}q_2 + \frac{2}{8}q_3$$

Berat badan = Average

$$= \frac{4}{8}(0.81) + \frac{2}{8}(0) + \frac{2}{8} = 0.40$$

$$q_2 = -\frac{0}{2}\log_2\frac{0}{2} - \frac{2}{2}\log_2\frac{2}{2} = 0$$

Berat badan = Underweight

$$q_3 = -\frac{0}{2} \log_2 \frac{0}{2} - \frac{2}{2} \log_2 \frac{2}{2} = 0$$

Berdasarkan perhitungan entropi diatas, maka diperoleh atribut berat badan sebagai node awal karena memiliki entropi terkecil.

#### Mengubah Tree Menjadi Rule

Dari rule yang sudah dihasilkan, maka diperoleh rule sebagai berikut:

- i. Rule 1 : IF berat = average OR berat = underweight THEN hipertensi = tidak.
- ii. Rule 2 : IF berat = overweight AND kelamin = wanita THEN hipertensi = ya.
- iii. Rule 3 : IF berat = overweight AND kelamin = pria THEN usia = muda THEN hipertensi = ya.
- iv. Rule 4 : IF berat = overweight AND kelamin = pria AND usia = tua THEN hipertensi = tidak.

#### Menguji Rule Untuk Prediksi

Nama	Usia	Berat	Kelamin	Hipertensi	Prediksi
Ali	Muda	Overweight	Pria	Ya	Ya
Edi	Muda	Underweight	Pria	Tidak	Tidak
Annie	Muda	Average	Wanita	Tidak	Tidak
Budiman	Tua	Overweight	Pria	Tidak	Tidak
Herman	Tua	Overweight	Pria	<b>Ya</b>	<b>Tidak</b>
Diji	Muda	Underweight	Pria	Tidak	Tidak
Rina	Tua	Overweight	Wanita	Ya	Ya
Gatot	Tua	Average	Pria	Tidak	Tidak

$$\text{Kesalahan (Error)} = \frac{1}{8} \times 100 \% = 12.5 \%$$

$$\text{Akurasi prediksi} = \frac{7}{8} \times 100 \% = 87.5 \%$$

### 3. METODOLOGI PENELITIAN

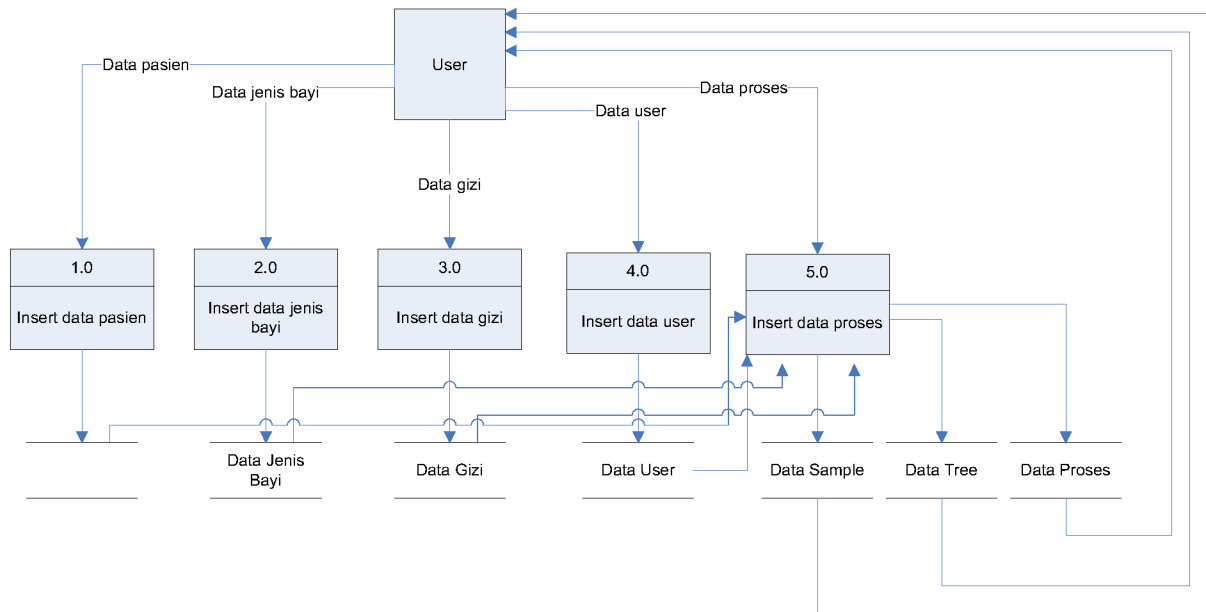
Pengembangan data mining:

- 3.1. Data Seleksi
- 3.2. Praproses
- 3.3. Transformasi
- 3.4. Pengujian Sistem
- 3.5. Interpretasi dan Evaluasi

### 4. HASIL DAN PEMBAHASAN

#### 4.1. Pemodelan Proses

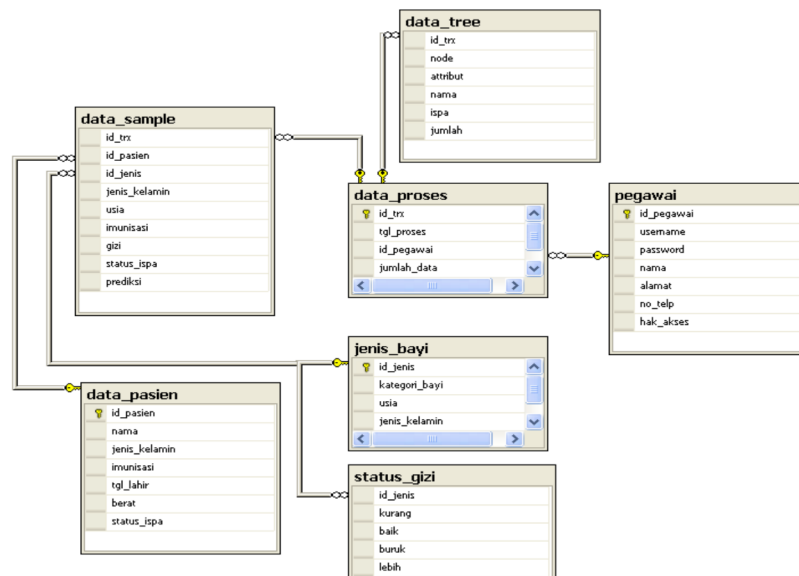
*Data Flow Diagram* (DFD) menggambarkan aliran data dalam sebuah system, dfd menggambarkan semua proses, meskipun proses tersebut terjadi dalam waktu yang berbeda, dan juga tidak ada loop ataupun cabang dalam DFD.



Gambar DFD Level 1

#### 4.2. Perancangan Tabel

Berdasarkan data-data yang diperoleh maka, dapat dibuat basis data yang saling berelasi. Pada penelitian ini, basis data yang diperlukan terdiri dari tujuh tabel, yaitu tabel data sampel, data tree, data proses, pegawai, data pasien, jenis bayi, dan status gizi.



Gambar Perancangan Tabel Yang di Bentuk

### 4.3. Implementasi Program

#### 4.3.1. Form Kategori Bayi

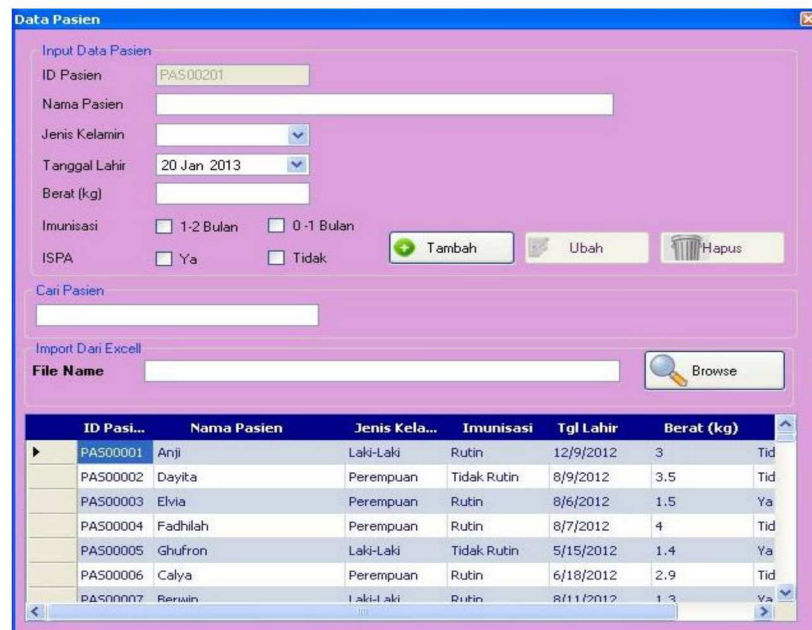
Form input data kategori bayi



ID Jenis	Kategori B...	Usia	Jenis Kela...	Buruk	Kura...	Bai
JEN001	USIA A	0-1	Laki-Laki	0-1.5	1.6-2.5	2.6-4
JEN002	USIA A	0-1	Perempuan	0-1.5	1.6-2.5	2.6-4
JEN003	USIA B	1-4	Laki-Laki	0-1.5	1.6-2.5	2.6-4
JEN004	USIA B	1-4	Perempuan	0-1.5	1.6-2.5	2.6-4
JEN005	USIA C	4-8	Laki-Laki	0-1.5	1.6-2.5	2.6-4

#### 4.3.2. Form Data Pasien


Form data pasien digunakan untuk input data pasien



ID Pasi...	Nama Pasien	Jenis Kela...	Imunisasi	Tgl Lahir	Berat (kg)
PAS00001	Anji	Laki-Laki	Rutin	12/9/2012	3
PAS00002	Dayita	Perempuan	Tidak Rutin	8/9/2012	3.5
PAS00003	Elvia	Perempuan	Rutin	8/6/2012	1.5
PAS00004	Fadhilah	Perempuan	Rutin	8/7/2012	4
PAS00005	Ghufron	Laki-Laki	Tidak Rutin	5/15/2012	1.4
PAS00006	Calya	Perempuan	Rutin	6/18/2012	2.9
PAS00007	Rarwin	Laki-Laki	Rutin	8/11/2012	1.3

#### 4.3.3. Form Decision Tree

Form decision tree digunakan untuk melihat hasil dari proses pohon keputusan.



No	Nama	Jenis K...	Kategori	Usia	Status Im...	Status Gizi	ISPA	Prediksi
1	Anji	Laki-Laki	USIA B	1	Rutin	Gizi Baik	Tidak	
2	Dayita	Peremp...	USIA C	5	Tidak Rutin	Gizi Baik	Tidak	
3	Elvia	Peremp...	USIA C	5	Rutin	Gizi Buruk	Ya	
4	Fadhilah	Peremp...	USIA C	5	Rutin	Gizi Baik	Tidak	
5	Ghufron	Laki-Laki	USIA D	8	Tidak Rutin	Gizi Buruk	Ya	
6	Calya	Peremp...	USIA C	7	Rutin	Gizi Baik	Tidak	
7	Berwin	Laki-Laki	USIA C	5	Rutin	Gizi Buruk	Ya	
8	Abdiel	Laki-Laki	USIA D	9	Rutin	Gizi Baik	Tidak	
9	Armila	Peremp...	USIA C	7	Rutin	Gizi Baik	Tidak	
10	Arsa	Laki-Laki	USIA B	1	Rutin	Gizi Baik	Tidak	
11	Dimas	Laki-Laki	USIA D	9	Rutin	Gizi Baik	Tidak	
12	Dzaky	Laki-Laki	USIA D	10	Tidak Rutin	Gizi Kurang	Ya	

#### 4.3.4. Form Pembentukan Rule

Digunakan untuk mengetahui rule yang dihasilkan dari pohon keputusan



**Rule :**

- 1: IF Gizi=Gizi Buruk THEN ISPA=Ya
- 2: IF Gizi=Gizi Lebih THEN ISPA=Tidak
- 3: IF Gizi=Gizi Baik AND Usia=USIA C THEN ISPA=Ya
- 4: IF Gizi=Gizi Baik AND Usia=USIA B AND Imunisasi=Tidak Rutin THEN ISPA=Ya
- 5: IF Gizi=Gizi Baik AND Usia=USIA B AND Imunisasi=Rutin THEN ISPA=Tidak
- 6: IF Gizi=Gizi Baik AND Usia=USIA D AND Imunisasi=Tidak Rutin THEN ISPA=Ya
- 7: IF Gizi=Gizi Baik AND Usia=USIA D AND Imunisasi=Rutin THEN ISPA=Tidak
- 8: IF Gizi=Gizi Kurang AND Usia=USIA C THEN ISPA=Ya
- 9: IF Gizi=Gizi Kurang AND Usia=USIA B AND Imunisasi=Tidak Rutin THEN ISPA=Ya
- 10: IF Gizi=Gizi Kurang AND Usia=USIA B AND Imunisasi=Rutin THEN ISPA=Tidak
- 11: IF Gizi=Gizi Kurang AND Usia=USIA D AND Imunisasi=Tidak Rutin THEN ISPA=Ya
- 12: IF Gizi=Gizi Kurang AND Usia=USIA D AND Imunisasi=Rutin THEN ISPA=Tidak

4.4. Berdasarkan hasil dari uji chi square pada program diatas adalah atribut usia dengan nilai kepercayaan 0.05 dan  $x^2$  hitung adalah 10.7,  $x^2$  tabel adalah 7.81 maka  $x^2$  hitung >  $x^2$  tabel yang artinya atribut usia mempengaruhi seseorang terkena ispa.

4.5. Atribut jenis kelamin dengan nilai kepercayaan 0.05 dan  $x^2$  hitung adalah 0 dan  $x^2$  tabel adalah 3.8 maka  $x^2$  hitung <  $x^2$  tabel yang artinya atribut jenis kelamin tidak mempengaruhi seseorang terkena ispa.



4.6. Atribut kriteria gizi dengan nilai kepercayaan 0.05 dan  $x^2$  hitung adalah 240.9 dan  $x^2$  tabel adalah 7.81 maka  $x^2$  hitung  $>$   $x^2$  tabel yang artinya atribut kriteria gizi mempengaruhi seseorang terkena ispa.

4.7. Atribut imunisasi dengan nilai kepercayaan 0.05 dan  $x^2$  hitung adalah 18.7 dan  $x^2$  tabel 3.8 maka  $x^2$  hitung  $>$   $x^2$  tabel yang artinya atribut imunisasi mempengaruhi seseorang terkena ispa.

## 5. KESIMPULAN DAN SARAN

5.1. Dari penelitian yang telah dilakukan, dihasilkan sebuah perangkat lunak (*software*) baru tentang penerapan data mining untuk klasifikasi prediksi penyakit ispa (*infeksi saluran pernapasan akut*) dengan menggunakan algoritma decision tree.

5.2. Perangkat lunak yang dihasilkan mampu menerapkan proses data mining untuk mengklasifikasikan penyakit ispa dengan atribut yang paling berpengaruh yang pertama adalah gizi sebesar 240.9 kemudian atribut yang berpengaruh kedua adalah imunisasi sebesar 18.7 dan yang terakhir adalah atribut usia sebesar 10.7. Dengan demikian, atribut yang paling tidak terpengaruh seseorang terkena ispa adalah jenis kelamin.

5.3. Saran yang berkaitan dengan system untuk menerapkan data mining dalam memprediksikan penyakit adalah aplikasi yang telah dibuat masih bisa dikembangkan lagi bagi yang tertarik dan berminat.

## DAFTAR PUSTAKA

- [1] Wahyudin. 2009. *Metode Iterative Dichotomizer 3 (ID3) Untuk Penyeleksian Penerimaan Mahasiswa Baru*. Program Pendidikan Ilmu Komputer. Universitas Pendidikan Indonesia.
- [2] Prawitasari Anita. *Klasifikasi Prediksi Penyakit DBD dan Tipes dengan Algoritma Decision Tree*. 2011. Jurusan system informasi. Universitas Gunadharma.
- [3] Berry, Michael J.A. dan Gordon S. Linoff. 2004. *Data mining techniques for marketing, sales, customer relationship management*. Second edition. Wiley publishing, Inc.
- [4] Kusriani dan Emha Taufiq Lutfi. 2009. *Algoritma Data Mining*. Yogyakarta; Andi Offset.
- [5] Gambetta, Windy. 2012. *Pohon Keputusan (Decision Tree)*. Departemen Teknik Informatika. Institute Teknologi Bandung. Bandung.