

# PREDIKSI PENYAKIT DEMAM BERDARAH DI PUSKESMAS NGEEMPLAK SIMONGAN MENGGUNAKAN ALGORITMA C4.5

<sup>1</sup> Saifur Rohman Cholil, <sup>2</sup> Aditya Febri Dwijayanto, <sup>3</sup> Tria Ardianita

<sup>1,3</sup> Program Studi Sistem Informasi, Fakultas Teknologi Informasi dan Komunikasi

<sup>2</sup> Program Studi Teknik Informatika, Fakultas Teknologi Informasi dan Komunikasi  
Universitas Semarang (USM)

Jl. Soekarno Hatta, Tlogosari Kulon, Kec. Pedurungan, Semarang, Indonesia 59160

E-mail : [cholil@usm.ac.id](mailto:cholil@usm.ac.id), [adityafebri3@gmail.com](mailto:adityafebri3@gmail.com), [triaardianita@gmail.com](mailto:triaardianita@gmail.com)

(Diterima: 16 Juni 2020 ,direvisi: 3 Juli 2020, disetujui: 18 Juli 2020)

## ABSTRACT

*Dengue Hemorrhagic Fever (DHF) is a disease whose main cause is the flaviviridae virus. This virus can be transmitted through mosquito bites. The spread of this disease is faster in urban areas than in rural areas due to the high population density. Aedes aegypti mosquito is very easy to spread dengue virus from one person to another because it has a domestic nature. The Ministry of Health has collaborated with local health centers in the DHF prevention program. Ngemplak Simongan Health Center is one of the public health centers located in the District of West Semarang that serves a variety of treatments for this type of disease, one of which is a patient with Dengue Hemorrhagic Fever (DHF). C4.5 algorithm is used to predict dengue fever which aims to produce a decision tree. The choice of using Algoritma is because it is widely used to describe a pattern / knowledge / information in the form of a decision tree explicitly. Application created using PHP programming language that produces prediction of dengue fever. The test results obtained an accuracy value of 94.44% so that the application program built can be used correctly.*

**Keywords:** *c4.5 algorithm, decision tree , dengue hemorrhagic fever*

## ABSTRAK

Demam Berdarah Dengue (DBD) adalah sebuah penyakit yang penyebab utamanya adalah virus *flaviviridae*. Virus ini dapat ditularkan melalui gigitan nyamuk. Penyebaran penyakit ini lebih cepat di area perkotaan dibandingkan di area pedesaan karena faktor tingginya kepadatan penduduk. Nyamuk *Aedes aegypti* sangat mudah menyebarkan virus *dengue* dari orang satu ke orang lain karena memiliki sifat domestik. Departemen kesehatan telah melakukan kerja sama dengan puskesmas sekitar dalam program penanggulangan penyakit DBD. Puskesmas Ngemplak Simongan merupakan salah satu puskesmas yang berada di Kecamatan Semarang Barat yang melayani berbagai macam pengobatan jenis penyakit, salah satunya adalah penderita Demam Berdarah Dengue (DBD). Algoritma C4.5 digunakan untuk prediksi penyakit demam berdarah yang bertujuan menghasilkan sebuah pohon keputusan. Pemilihan penggunaan Algoritma ini karena banyak digunakan untuk menggambarkan suatu pola/pengetahuan/informasi dalam bentuk pohon keputusan secara eksplisit. Aplikasi yang dibuat menggunakan bahasa pemrograman PHP yang menghasilkan prediksi penyakit demam berdarah. Hasil pengujian didapatkan nilai akurasi sebesar 94.44% sehingga aplikasi program yang dibangun dapat digunakan secara benar.

**Kata Kunci:** algoritma c4.5, pohon keputusan, demam berdarah dengue

## 1 PENDAHULUAN

Pesatnya perkembangan ilmu pengetahuan dan teknologi khususnya di bidang teknologi informasi dapat memunculkan suatu inovasi baru untuk menyajikan dan mengelola suatu informasi demi memenuhi kebutuhan informasi. Peranan teknologi pada bidang kesehatan maupun kedokteran telah banyak membantu menolong jiwa manusia dan telah menunjukkan peranan pentingnya. Komputer telah banyak digunakan untuk menganalisis bagian dalam organ tubuh manusia yang sulit dilihat, mencari tahu obat yang tepat, bahkan untuk mendiagnosis penyakit.

Demam Berdarah Dengue (DBD) adalah penyakit yang ditularkan oleh nyamuk lewat gigitannya yang mengandung virus *flaviviridae* serta dapat menyebabkan kematian. Penyebaran penyakit DBD di area perkotaan lebih cepat dibandingkan di area pedesaan karena faktor tingginya kepadatan penduduk. Nyamuk *Aedes aegypti* sangat mudah menyebarkan virus *dengue* dari orang satu ke orang lain karena memiliki sifat domestik. Penanganan program penyakit DBD yang dilakukan oleh pemerintah khususnya Departemen Kesehatan telah melakukan kerja sama dengan semua puskesmas sekitar untuk deteksi dan pengobatan penyakit tersebut.

Puskesmas Ngemplak Simongan merupakan salah satu puskesmas yang berada di kecamatan Semarang Barat yang melayani berbagai macam pengobatan jenis penyakit salah satunya adalah penderita Demam Berdarah Dengue (DBD). Puskesmas Ngemplak Simongan merupakan puskesmas yang jumlah pasien penderita DBD terbanyak di kecamatan Semarang Barat. Dimana dalam melakukan pengecekan pasien yang terkena penyakit demam berdarah harus melalui diagnosa terlebih dahulu. Kemudian dilanjutkan dengan pemeriksaan laboratorium jika diperlukan oleh pasien. Tujuan penelitian ini adalah memprediksi penyakit demam berdarah menggunakan algoritma C4.5 dan membuat sistem berbasis web untuk melakukan prediksi penyakit tersebut. Prediksi dilakukan dengan cara mengambil data-data yang sudah ada pada Puskesmas tersebut tentang penyakit DBD yang selanjutnya disebut sebagai data training.

Algoritma C4.5 telah banyak digunakan untuk memprediksi penyakit, diantaranya dari penelitian [1] melakukan prediksi penyakit *Rheumatoid Arthritis* (RA) dan mempunyai akurasi rata-rata 84%. Prediksi kelangsungan hidup kanker paru-paru [2] dengan metode SVM, Naïve Bayes dan C4.5, hasil akurasi menunjukkan algoritma C4.5 memiliki kinerja lebih baik dibandingkan yang lain dalam memprediksi kanker paru-paru. Algoritma C4.5 juga digunakan [3] untuk diagnosis turbin angin, deteksi dan diagnosis kesalahan sistem *photovoltaic* (PV) yang terhubung dengan jaringan [4] yang mempunyai akurasi pengujian sebesar 99,8%.

Pohon keputusan algoritma C4.5 digunakan untuk penalaran tegas manajemen tanggap darurat [5], konstruksi kendala otomatis untuk model *Mixed-Integer Linear Programming* (MILP) dari data [6] dan pengembangan pohon keputusan [7]. Algoritma ini juga digunakan oleh [8] untuk penilaian stabilitas tegangan online dan hasilnya bisa membantu operator sistem menilai status stabilitas tegangan secara *real-time*.

## 2 TINJAUAN PUSTAKA

### 2.1 Data Mining

Data *Mining* merupakan serangkaian proses untuk mencari informasi maupun pola yang berupa informasi dari suatu basis data dengan memanfaatkan teknik tertentu [9]. Hasil dari informasi diperoleh dari mengekstraksi dan menganalisis pola data yang ada dalam basis data [10]. *Knowledge Discovery Database (KDD)* merupakan metode yang saintifik dalam data *mining*, karena digunakan untuk mencari informasi maupun pengetahuan baru yang terdapat dalam kumpulan data. Data mining juga merupakan pembelajaran yang memiliki basis induksi karena terdapat proses pembentukan definisi dengan konsep umum dengan cara mengobservasi contoh yang spesifik dari semua konsep yang dipelajari [11]. Data mining sendiri telah banyak digunakan sebagai penelitian, salah satunya pada [12] yang digunakan untuk memprediksi masa studi mahasiswa dengan hasil nilai evaluasi penelitian yang lebih akurat dibandingkan dengan analisa secara manual. Penganalisaan data mining diperlukan sebuah software yang disebut dengan Rapid Miner. *Software* ini dibuat dengan menggunakan bahasa pemrograman Java serta dapat dijalankan pada sistem operasi apapun. Seperti pada penelitian [13] tentang analisa data hasil keuntungan dengan Rapid Miner yang dapat membantu proses analisis data dengan cepat dan efisien dengan kepastian kemungkinan sekitar 0,59 %.

### 2.2 Algoritma

Algoritma merupakan sistem kerja komputer yang digunakan untuk menyusun langkah yang sistematis agar dapat digunakan untuk memecahkan suatu masalah. Algoritma juga bersifat logis dan sistematis karena susunan langkahnya disusun secara tertulis dan berurutan. Algoritma sangat

berperan penting dalam dunia komputer karena merupakan salah satu pembangun dari software [14].

Algoritma juga memiliki prosedur yang jelas dalam menyelesaikan suatu persoalan karena terdapat langkah-langkah dan jumlahnya juga terbatas. Penyajian algoritma dapat berupa gambar maupun *pseudocode*. *Pseudocode* merupakan kode pemrograman yang hampir sama dengan Pascal, ataupun C, sehingga dapat digunakan untuk menggambarkan algoritma yang diinginkan oleh pemrogram. Untuk algoritma dalam bentuk gambar maka dapat disajikan dalam bentuk *flowchart*.

### 2.3 Algoritma C4.5

Algoritma C4.5 adalah suatu pengembangan induksi dari ID3 (*Iterative Dichotomiser 3*) karena C4.5 merupakan sebuah algoritma yang dapat digunakan untuk membuat sebuah pohon keputusan (*decision tree*). Pohon-pohon keputusan dibangun berdasarkan kriteria pembentuk suatu keputusan [15]. Pengembangan yang terdapat dalam C4.5 adalah dapat mengatasi nilai suatu atribut yang hilang (*missing value*), dapat mengolah data berjenis diskret maupun numerik, menghasilkan aturan yang mudah dalam pengintepretasian serta *prunning* atau pemangkasan [16].

Pemilihan atribut test untuk simpul, maka didasarkan pada atribut yang memiliki nilai *Gain Ratio* paling tinggi dari semua atribut yang ada. Tahapan untuk menghitung *GainRatio* dapat ditunjukkan menggunakan beberapa persamaan berikut [16] :

- a. Tahapan pertama yaitu menghitung nilai *Entropy* dengan menggunakan rumus yang ditunjukkan oleh rumus (1) berikut ini.

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \quad (1)$$

Keterangan :

S : himpunan kasus

n : jumlah partisi himpunan S

Pi : kasus partisi ke-i

- b. Tahap kedua yaitu mencari nilai *Gain* dengan menggunakan rumus yang telah ditunjukkan pada persamaan (2) berikut.

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (2)$$

Keterangan :

S : kasus

A : atribut

n : jumlah partisi atribut ke-A

|Si| : jumlah kasus partisi ke-i

|S| : jumlah kasus pada S

- c. Tahap ketiga yaitu menghitung *Split Info* dengan rumus yang ditunjukkan dalam persamaan (3) dibawah ini.

$$SplitInfo(S,A) = - \sum_{i=1}^n \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|} \quad (3)$$

Keterangan :

S = ruang sample data

A = atribut

Si = jumlah sample atribut ke-i

- d. Tahap keempat yaitu mencari nilai *Gain Ratio* dengan menggunakan rumus seperti tertera dalam persamaan (4) berikut ini.

$$Gain Ratio (S,A) = \frac{Gain(S,A)}{SplitInfo(S,A)} \quad (4)$$

Keterangan :

Gain (S,A) = informasi gain atribut A

SplitInfo (S,A) = split informasi atribut A

Jadi jika ada atribut yang memiliki nilai *Gain Ratio* tertinggi, maka akan dipilih menjadi *node* sebagai atribut test simpul.

### 2.4 Pohon Keputusan (*Decision Tree*)

Saifur Rohman Cholil dkk : Prediksi Penyakit Demam Berdarah Di Puskesmas Ngemplak Simongan Menggunakan Algoritma C4.5

Pohon keputusan merupakan suatu struktur yang mengklasifikasikan kumpulan data ke dalam tabel kelas yang telah ditentukan [17]. Tabel biasanya terdapat atribut yang berperan sebagai parameter dan record untuk membangun sebuah data. Manfaat menggunakan pohon keputusan adalah dapat mengubah sebuah fakta menjadi sebuah pohon keputusan yang lebih dapat dimengerti dan simpel dengan kombinasi bahasa alami. Pohon keputusan memiliki fungsi untuk menyelidiki data, mencari hubungan tersembunyi antara calon variabel target dengan variabel input yang akan digunakan. Perpaduan antara pemodelan dan eksplorasi data pada pohon keputusan sangat direkomendasikan sebagai langkah awal pemodelan serta model akhir dari beberapa teknik lain [7]. Simpul node dari *decision tree* pohon dibedakan menjadi tiga [11], yaitu :

- a. Simpul akar (*root node*) yang letaknya berada pada bagian teratas serta tidak memiliki inputan namun terkadang memiliki output lebih dari satu atau tidak sama sekali.
- b. Simpul percabangan (*internal node*) yang hanya mempunyai satu inputan dan dua outputan minimal.
- c. Simpul daun (*leaf node*) yang hanya mempunyai satu inputan dan sama sekali tidak memiliki outputan.

### 3 METODE PENELITIAN

Langkah yang digunakan pada penelitian ini untuk memprediksi penyakit DBD yang terdapat pada Puskesmas Ngemplak Simongan, yaitu dengan langkah KDD (*Knowledge Discovery in Databases*), yang telah terdapat beberapa tahapan yaitu seleksi, praproses, transformasi, data mining, interpretasi serta evaluasi.

#### 3.1 Jenis Data

Penelitian ini menggunakan jenis data yang terdiri dari data kuantitatif dan data kualitatif. Data kuantitatif merupakan jenis data yang dapat dihitung dan berbentuk angka maupun bilangan, sedangkan data kualitatif merupakan data yang mengandung informasi dan berbentuk kalimat. Semua data yang ada diperoleh berasal dari data primer karena data yang dikumpulkan berasal dari narasumber pertama, yaitu Kepala Puskesmas Ngemplak Simongan.

#### 3.2 Tahapan Metodologi Penelitian

##### a. Pengumpulan Data

Proses pengumpulan data merupakan mengumpulkan semua data yang diperlukan untuk mendukung proses algoritma C4.5. Semua data yang dibutuhkan dikumpulkan secara langsung dengan cara melakukan observasi di lapangan serta wawancara secara langsung dengan pihak terkait. Data primer yang digunakan adalah data yang berasal dari puskesmas yang berisi data pasien demam berdarah yang terdiri dari data training dan data testing.

##### b. Seleksi Data

Seleksi data merupakan proses menyortir atau memilih data-data yang akan dipergunakan dalam proses mining, namun tetap memperlihatkan bentuk data aslinya.

##### c. Transformasi Data

Transformasi data merupakan suatu cara untuk mengubah data asli ke bentuk lain agar dapat dilakukan pemrosesan dengan menggunakan algoritma C4.5.

##### d. Proses Perhitungan

Proses perhitungan untuk semua atribut maupun variable, *entropy* dapat menggunakan rumus (1), *information gain* dengan rumus (2), *splitinfo* menggunakan rumus (3), serta *gainratio* menggunakan rumus (4).

##### e. *Decision Tree* atau Pohon Keputusan

Pohon keputusan merupakan hasil dari perhitungan yang dilakukan sebelumnya, dengan dilakukannya perhitungan berulang kali hingga semua atribut memiliki kelas dan sudah tidak dapat lagi dilakukan perhitungan.

##### f. Pengujian

Tahap pengujian merupakan tahap pengetestan dimana hal ini dilajukan dengan tujuan untuk mengetahui semua proses dapat bekerja sesuai keinginan atau tidak.

g. Analisis Hasil Pengujian

Tujuan dilakukannya proses analisis adalah untuk memastikan hasil uji benar-benar dapat dipertanggungjawabkan. Proses analisis dilakukan dengan cara menghitung kembali hasil pengujian secara manual dicocokkan dengan hasil perhitungan sistem.

4 HASIL DAN PEMBAHASAN

4.1 Pengumpulan Data

Pengumpulan data dilakukan dengan cara meminta data ke Puskesmas Ngeplak Simongan. Data yang diperlukan adalah data pasien penderita Demam Berdarah di Puskesmas Ngeplak Simongan yang diambil dari Kepala Sub Bagian (Kasubag) Kesehatan Masyarakat (Kesmas) Puskesmas Ngeplak Simongan. Adapapun bentuk datanya seperti pada Gambar 1.

NO	NAMA	JENIS KELAMIN	USIA	SUHU TUBUH	BINTIK MERAH	MIMISAN	TROMBOSIT	DIAGNOSA
1	Anggita	PEREMPUAN	5	39°C	0	TIDAK	180000	TIDAK
2	Milhayla	PEREMPUAN	3,5	39°C	0	TIDAK	91000	YA
3	Ahlyan	LAKI - LAKI	19	39°C	4	YA	90531	YA
4	Reza	LAKI - LAKI	10	40°C	1	YA	90025	YA
5	Fetri	LAKI - LAKI	8	38°C	3	TIDAK	170000	TIDAK
6	Wistiani	PEREMPUAN	34	39°C	1	YA	88000	YA
7	Siva	PEREMPUAN	25	39°C	0	TIDAK	73000	YA
8	Kharisma	PEREMPUAN	27	41°C	1	YA	80000	YA
9	Ahni Nur Shodiq	LAKI - LAKI	8	39°C	1	TIDAK	92000	YA
10	Celine Christa	PEREMPUAN	5,5	38°C	0	YA	150345	TIDAK
11	Mavidatul	PEREMPUAN	20	39°C	0	YA	90500	YA
12	Iqbal Ramadhan	LAKI - LAKI	18	39°C	1	TIDAK	79000	YA
13	Aviva Aulaningtyas	PEREMPUAN	4,5	39°C	0	TIDAK	99000	YA
14	Fadilla Choirunisa	PEREMPUAN	15	39°C	2	TIDAK	76062	YA

Gambar 1. Data Penyakit Demam Berdarah

4.2 Seleksi Data

Proses seleksi data dilakukan dengan cara memilih atribut atau variabel yang dibutuhkan. Pemilihan atribut dilakukan dengan pertimbangan atribut yang dipilih sudah sesuai dengan atribut yang dibutuhkan untuk melakukan proses prediksi demam berdarah. Dari data yang tersedia diambil beberapa atribut antara lain: suhu tubuh, bintik merah, mimisan, trombosit dan diagnosa.

4.3 Transformasi Data

Transformasi data dilakukan dengan cara mengubah data yang di dapat dari Puskesmas Ngeplak Simongan ke dalam format excel atau format CSV. Langkah ini dilakukan untuk analisis dan visualisasi data dengan aplikasi *rapidminer*.

4.4 Proses Perhitungan Manual Metode Algoritma C4.5

Adapun langkah perhitungan manual dengan menggunakan metode Algoritma C4.5 adalah mencari nilai dari *entropy*, *gain*, *split info*, dan *gain ratio*. Data yang diperoleh dari Kepala Sub Bagian (Kasubag) Kesehatan Masyarakat (Kesmas) Puskesmas Ngeplak Simongan yang telah dilakukan proses seleksi dan transformasi data seperti pada Tabel 1.

Langkah pertama yaitu menghitung nilai *entropy*. Perhitungan nilai *entropy* dilakukan sesuai dengan rumus persamaan 1 dan proses perhitungannya sebagai berikut :

$$Entropy = \left(-\frac{28}{101} \times \log_2 \frac{28}{101}\right) + \left(-\frac{73}{101} \times \log_2 \frac{73}{101}\right) = 0,8516$$

$$Entropy(38^\circ C) = -\left(\frac{8}{26}\right) * \log_2 \left(\frac{8}{26}\right) - \left(\frac{18}{26}\right) * \log_2 \left(\frac{18}{26}\right) = 0,8905$$

$$Entropy(39^\circ C) = -\left(\frac{50}{59}\right) * \log_2 \left(\frac{50}{59}\right) - \left(\frac{9}{59}\right) * \log_2 \left(\frac{9}{59}\right) = 0.6162$$

$$Entropy(40^\circ C) = -\left(\frac{12}{13}\right) * \log_2 \left(\frac{12}{13}\right) - \left(\frac{1}{13}\right) * \log_2 \left(\frac{1}{13}\right) = 0.3912$$

$$Entropy(41^\circ C) = -\left(\frac{3}{3}\right) * \log_2 \left(\frac{3}{3}\right) - \left(\frac{0}{3}\right) * \log_2 \left(\frac{0}{3}\right) = 0$$

Analisa penyelesaian 1 node yaitu menghitung nilai total keputusan yang dinyatakan Ya (73) dan dinyatakan Tidak (28), sedangkan untuk jumlah 101 merupakan total dari seluruh data. Langkah kedua yaitu menghitung nilai *gain*. Perhitungan ini sesuai dengan rumus persamaan 2 dan proses perhitungannya sebagai berikut :

$$Gain = 0,8516 - \left(\left(\frac{26}{101} * 0,8905\right) + \left(\frac{59}{101} * 0,6162\right) + \left(\frac{13}{101} * 0,3912\right) + \left(\frac{3}{101} * 0\right)\right) = 0,2121$$

Langkah ketiga yaitu menghitung nilai *split info*. Perhitungan ini sesuai dengan rumus persamaan 3 dan proses perhitungannya sebagai berikut :

$$Split Info = -\frac{26}{101} * \log_2 \left(\frac{26}{101}\right) - \frac{59}{101} * \log_2 \left(\frac{59}{101}\right) - \frac{13}{101} * \log_2 \left(\frac{13}{101}\right) - \frac{3}{101} * \log_2 \left(\frac{3}{101}\right) = 1.4884$$

Langkah keempat yaitu menghitung nilai *gain ratio*. Perhitungan ini sesuai dengan rumus persamaan 4 dan proses perhitungannya sebagai berikut :

$$Gain Ratio = \frac{0,2121}{1,4884} = 0,1425$$

**Tabel 1. Data Training**

Suhu Tubuh	Bintik Merah	Mimisan	Trombosit	Diagnosa
39°C	0	Tidak	resiko rendah	tidak
39°C	0	Tidak	resiko sedang	ya
39°C	4	Ya	resiko tinggi	ya
40°C	1	Ya	resiko tinggi	ya
38°C	3	Tidak	resiko rendah	tidak
39°C	1	Ya	resiko tinggi	ya
39°C	0	Tidak	resiko tinggi	ya
41°C	1	Ya	resiko tinggi	ya
39°C	1	Tidak	resiko sedang	ya
38°C	0	Ya	resiko rendah	tidak
39°C	0	Ya	resiko tinggi	ya
39°C	0	Tidak	resiko rendah	tidak
39°C	0	Tidak	resiko sedang	ya
39°C	4	Ya	resiko tinggi	ya
40°C	1	Ya	resiko tinggi	ya
38°C	3	Tidak	resiko rendah	tidak
39°C	1	Ya	resiko tinggi	ya
39°C	0	Tidak	resiko tinggi	ya
41°C	1	Ya	resiko tinggi	ya
39°C	1	Tidak	resiko tinggi	ya
39°C	0	Tidak	resiko sedang	ya
39°C	2	Tidak	resiko tinggi	ya
39°C	0	Tidak	resiko tinggi	ya
40°C	2	Ya	resiko tinggi	ya
40°C	0	Ya	resiko tinggi	ya
39°C	0	Tidak	resiko rendah	tidak
40°C	1	Ya	resiko tinggi	ya
39°C	1	Ya	resiko tinggi	ya
39°C	0	Tidak	resiko sedang	ya
39°C	2	Tidak	resiko tinggi	ya
38°C	1	Ya	resiko rendah	tidak
39°C	0	Tidak	resiko sedang	ya
41°C	3	Ya	resiko tinggi	ya
39°C	1	Ya	resiko tinggi	ya
39°C	0	Ya	resiko tinggi	ya
39°C	4	Tidak	resiko rendah	tidak
38°C	0	Tidak	resiko rendah	tidak
39°C	0	Tidak	resiko sedang	ya
38°C	4	Ya	resiko tinggi	ya
38°C	7	Ya	resiko tinggi	ya
39°C	0	Ya	resiko sedang	ya
39°C	5	Tidak	resiko tinggi	ya
39°C	0	Tidak	resiko sedang	ya
39°C	0	Ya	resiko sedang	ya
38°C	0	Tidak	resiko sedang	tidak
41°C	2	Ya	resiko tinggi	ya
39°C	1	Ya	resiko tinggi	ya
39°C	0	Tidak	resiko sedang	ya
39°C	0	Tidak	resiko rendah	tidak
39°C	1	Ya	resiko sedang	ya

Perhitungan suhu tubuh, bintik merah, mimisan dilakukan proses perhitungan seperti di atas dan hasil semua perhitungan seperti ditunjukkan pada Tabel 2.

**Tabel 2. Penghitungan Node 1**

Node	Attribut	Total data	Ya	Tidak	Entropy	Gain	Split info	Gain ratio
1	Diagnosa	101	73	28	0,8516	0,2121	1,4884	0,1425
	Suhu tubuh							
	38°C	26	8	18	0,8905			
	39°C	59	50	9	0,6162			
	40°C	13	12	1	0,3912			
	41°C	3	3	0	0			
						0,06781	1,8461	0,0367
	Bintik merah							
	0	56	35	21	0,9544			
	1	23	21	2	0,4262			
	2	9	7	2	0,7642			
	3	4	3	1	0,8113			
	4	6	4	2	0,9183			
	5	2	2	0	0			
	7	1	1	0	0			
						0,0465	0,9982	0,0466
	Mimisan							
	Ya	53	44	9	0,6573			
	Tidak	48	29	19	0,9685			
						0,6706	1,5526	0,4319
	Trombosit							
	Resiko tinggi	40	40	0	0			
	Resiko sedang	37	33	4	0,4942			
	Resiko rendah	24	0	4	0			

Hasil dari Tabel 2, dapat diketahui bahwa nilai atribut dengan *GainRatio* paling tinggi adalah Saifur Rohman Cholil dkk : *Prediksi Penyakit Demam Berdarah Di Puskesmas Ngemplak Simongan Menggunakan Algoritma C4.5*

**Trombosit**, yaitu sebesar 0,4319. Jadi dengan demikian **Trombosit** menjadi *node* akar. Adapun tiga atribut dari **Trombosit** yaitu **Resiko Tinggi**, **Resiko Sedang** dan **Resiko Rendah**. Dari ketiga nilai atribut tersebut, nilai **Resiko Tinggi** dan **Resiko Rendah** telah mengklasifikasikan permasalahan menjadi satu yaitu keputusan **Tidak**, sehingga **tidak perlu** dilakukan penghitungan lebih lanjut, namun untuk hasil **nilai atribut Resiko Sedang** masih perlu dilakukan penghitungan ulang. Perhitungan node ditunjukkan oleh Tabel 3 di bawah ini.

**Tabel 3. Penghitungan Node 1.1**

Node	Atribut	Total data	Ya	Tidak	Entropy	Gain	Split info	Gain ratio
1.1	Trombosit resiko sedang	37	33	4	0,4942			
						0,1382	0,8707	0,1587
	Suhu tubuh	38°C	7	3	4	0,9852		
		39°C	29	28	1	0,2164		
		40°C	1	1	0	0		
		41°C	0	0	0	0		
						0,0656	1,0331	0,0635
	Bintik merah	0	25	21	4	0,6243		
		1	10	10	0	0		
		2	0	0	0	0		
		3	1	1	0	0		
		4	0	0	0	0		
		5	0	0	0	0		
		7	0	0	0	0		
						0,1077	1,0257	0,1050
	Mimisan	Ya	17	16	1	0,3228		
		Tidak	10	7	3	0,8813		

Jika dilihat dari Tabel 3, maka dapat disimpulkan bahwa nilai atribut dengan *GainRatio* paling tinggi adalah **Suhu Tubuh**, yaitu sebesar 0,1587. Maka, **Suhu Tubuh** dapat dijadikan sebagai *node* cabang dari nilai atribut sebesar 38°C atau **Sedang**. Terdapat empat nilai atribut dari **Suhu Tubuh**, yaitu 38°C, 39°C, 40°C dan 41°C. Dari keempat nilai atribut tersebut, nilai atribut 39°C, 40°C dan 41°C atau **Tinggi** telah digolongkan kasus menjadi 1, yaitu keputusan **YA**, dan nilai atribut **Tinggi** sudah mengklasifikasikan kasus menjadi 1, yaitu keputusan **YA**, maka tidak perlu dilakukan penghitungan ke tahap selanjutnya, namun untuk nilai atribut 38°C atau **Sedang** masih perlu dilakukan penghitungan lagi. Perhitungan node berikutnya ditunjukkan oleh Tabel 4.

**Tabel 4. Penghitungan Node 1.1.2**

Node	Atribut	Total data	Ya	Tidak	Entropy	Gain	Split info	Gain ratio
1.1.2	Trombosit resiko sedang dan Suhu Tubuh 38°C	7	3	4	0,9852			
						0,2917	0,7478	0,3901
	Bintik merah	0	5	2	3	0,9710		
		1	0	0	0	0		
		2	1	1	0	0		
		3	1	1	0	0		
		4	0	0	0	0		
		5	0	0	0	0		
		7	0	0	0	0		
						0,1281	0,9852	0,1300
	Mimisan	Ya	4	3	1	0,8183		
		Tidak	3	1	2	0,9183		

Kesimpulan yang dapat diambil dari Table 4 adalah bahwa atribut dengan nilai *GainRatio* paling tinggi adalah **Bintik Merah**, yaitu sebesar 0,3901. Sehingga **Bintik Merah** dapat dijadikan sebagai *node* cabang dari nilai atribut **Suhu Tubuh**. Terdapat empat nilai atribut dari **Bintik Merah**, 0,1,2,3,4,5 dan 7. Dari keenam nilai atribut tersebut nilai atribut **1,2,3,4,5,7** telah mengklasifikasikan suatu kasus menjadi 1, yaitu keputusan **YA**, namun untuk nilai atribut **0** masih diperlukan penghitungan ulang kembali. Perhitungan node berikutnya ditunjukkan oleh Tabel 5.

**Tabel 5. Penghitungan Node 1.1.3**

Node	Atribut	Total data	Ya	Tidak	Entropy	Gain	Split info	Gain ratio
1.1.2	Trombosit Resiko Sedang dan Suhu Tubuh 38°C dan Bintik Merah "0"	5	2	3	0,99710			
						0,4200	0,9710	0,4325
	Mimisan	Ya	3	2	1	0,9183		
		Tidak	2	0	2	0		

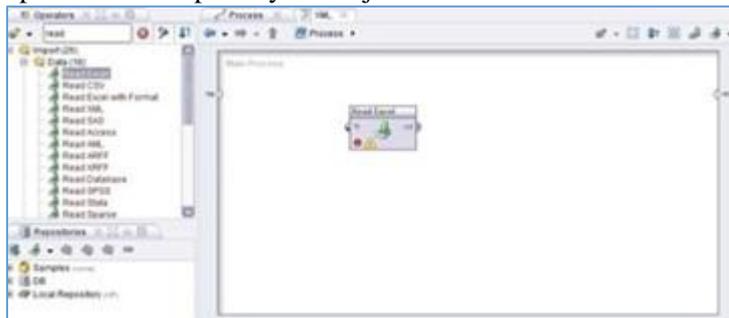
Dari Tabel 5, telah diketahui bahwa atribut dengan **GainRatio** dari **Mimisan**, yaitu sebesar 0,4325 dan nilai opsi atribut dari **Mimisan**, yaitu **YA** dan **TIDAK**. Dari kedua jumlah nilai atribut tersebut, maka nilai atribut **TIDAK** telah digolongkan kasus menjadi 1, yaitu berupa keputusan **Tidak**, serta nilai untuk nilai atribut **YA** telah menggolongkan kasus menjadi 1 dengan keputusan **Ya**, jadi tidak diperlukan penghitungan ke tahap selanjutnya untuk jenis atribut tersebut.

**4.5 Desain Sistem**

**Input Data Training**, dalam tahap ini software yang digunakan adalah Rapid Miner. Langkah- langkahnya antara lain sebagai berikut :

**a. Read Excel**

Read excel digunakan untuk membaca data dari format microsoft excel. Data yang sudah ada ke dalam aplikasi rapidminer. Tampilannya ditunjukkan oleh Gambar 2.

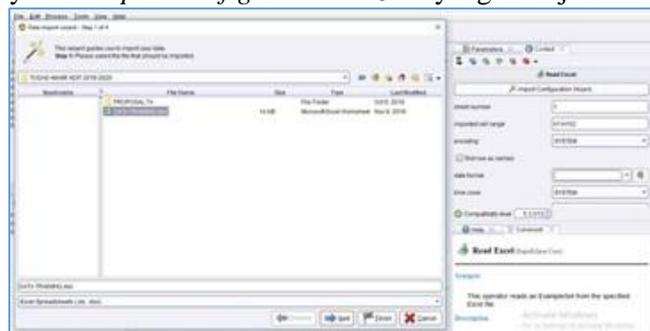


**Gambar 2. Read Excel**

Pada Gambar 2 di atas dijelaskan tentang langkah pertama pada desain sistem menggunakan aplikasi *rapidminer*, pada tahap ini menambahkan operator *read excel* yang digunakan untuk membaca data dari file microsoft excel.

**b. Import Configuration Wizard**

Tahapan berikutnya ada *import configuration wizard* yang ditunjukkan oleh Gambar 3.

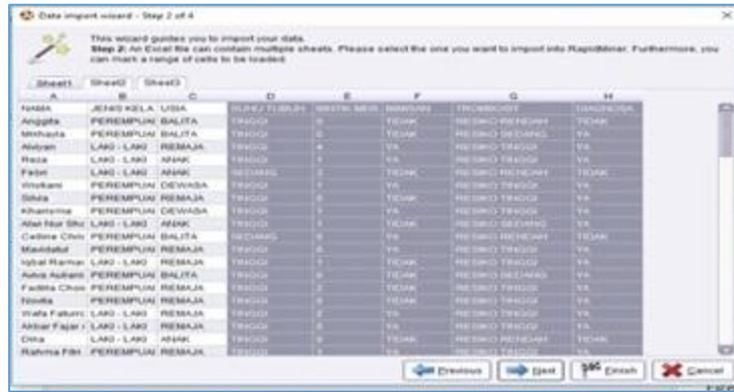


**Gambar 3. Tampilan Import Configuraion Wizard**

Gambar 3 menjelaskan tentang cara mengimport file excel untuk disesuaikan formatnya dengan format rapidminer. Pada bagian sebelah kanan kemudian pilih file *dataset* yang sudah siap kemudian pilih Next.

**c. Blok kolom yang diperlukan**

Langkah ketiga adalah memblok kolom yang akan digunakan dalam pembentukan model seperti ditunjukkan oleh Gambar 4.

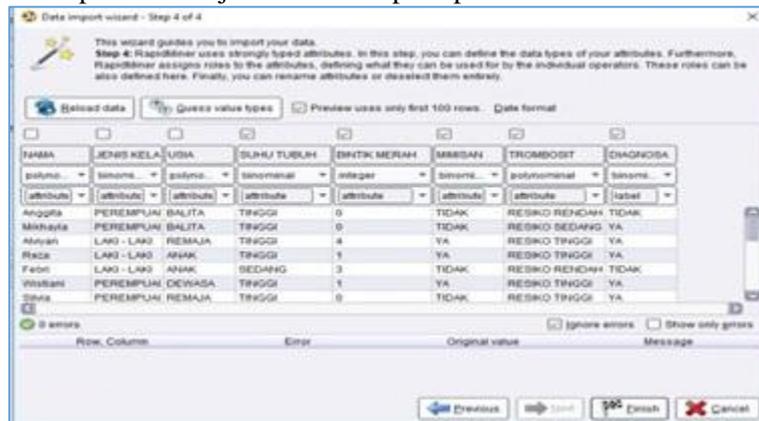


**Gambar 4. Blok Kolom Yang Diperlukan**

Gambar 4 di atas menjelaskan tentang memilih kolom yang diperlukan untuk keperluan analisis data.

**d. Penentuan jenis atribut**

Berikutnya adalah penentuan jenis atribut seperti pada Gambar 5 di bawah ini.



**Gambar 5. Penentuan Jenis Atribut**

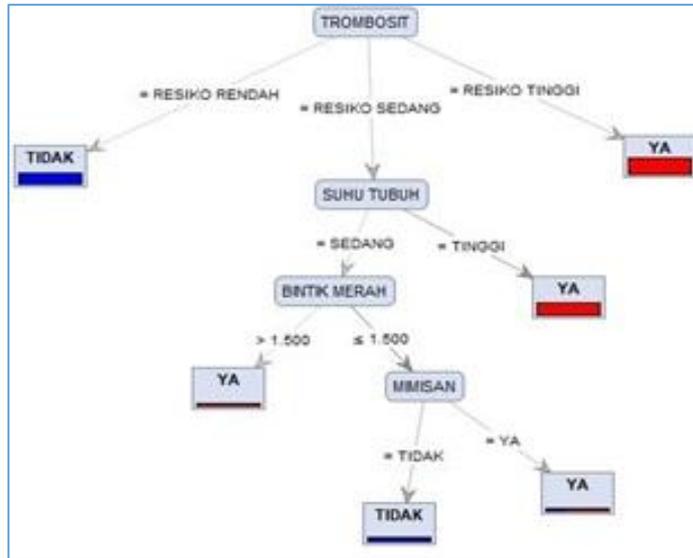
Gambar 5 di atas menjelaskan tentang penentuan jenis atribut. Pada tahap ini memilih atribut seperti pada pada penjelasan di bawah ini.

Tahap terakhir dalam *Read Excel* adalah penentuan jenis atribut deskripsinya sebagai berikut :

- (1) Suhu Tubuh = Binominal dan Atribut
- (2) Bintik Merah = Integer dan Atribut
- (3) Mimisan = Binominal dan Atribut
- (4) Trombosit = Polynomial dan Atribut
- (5) Diagnosa = Binominal dan Label

**4.6 Pembentukan Hasil Decision Tree**

Hasil dari perhitungan node didapatkan *decision tree* seperti Gambar 6. Gambar 6 di atas menjelaskan pohon keputusan yang didapatkan dari hasil perhitungan *node*. Tahap ini merupakan tahap penghitungan manual secara keseluruhan dan hasil nilai dari pengujian data set yang telah dilakukan melalui aplikasi *rapidminer*.



Gambar 6. Decision Tree

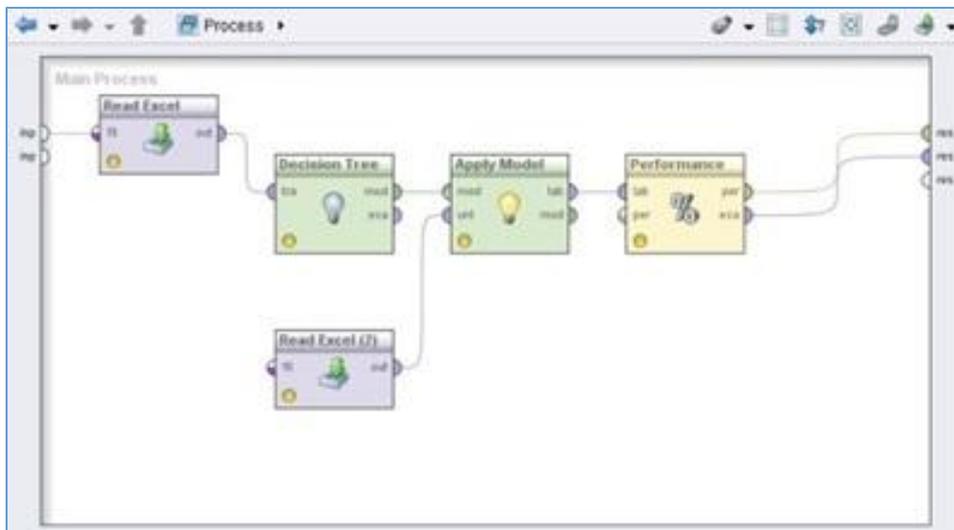
Selanjutnya hasil dari pohon keputusan diatas akan diekstraksi menjadi rule-rule seperti pada Tabel 6 berikut ini.

Tabel 6 Rule

Id	Rule
1	If Trombosit = Resiko Tinggi Then Diagnosa = YA
2	If Trombosit = Resiko Sedang And Suhu Tubuh = Tinggi Then Diagnosa = YA
3	If Trombosit = Resiko Sedang And Suhu Tubuh = Sedang And Bintik Merah = ≤ 1.500 And Mimisan = Ya Then Diagnosa = YA
4	If Trombosit = Resiko Sedang And Suhu Tubuh = Sedang And Bintik Merah = ≤ 1.500 And Mimisan = Tidak Then Diagnosa = TIDAK
5	If Trombosit = Resiko Sedang And Suhu Tubuh = Sedang And Bintik Merah = > 1.500 Then Diagnosa = YA
6	If Trombosit = Resiko Rendah Then Diagnosa = TIDAK

4.7 Pengujian Data dan Analisis Hasil Pengujian

Hasil dari perhitungan data testing ditunjukkan oleh Gambar 7 dan hasil akurasi ditunjukkan oleh Gambar 8.



Gambar 7. Pengujian Data Testing Pada Rapid Miner

	true YA	true TIDAK	class precision
pred. YA	23	0	100.00%
pred. TIDAK	2	11	84.62%
class recall	92.00%	100.00%	

**Gambar 8. Hasil Akurasi Pengujian Data Training Pada Rapid Miner**

Pada Gambar 7 di atas menunjukkan tahap pengujian data testing pada *rapidminer* bertujuan mengetahui hasil akurasi algoritma C4.5 pada prediksi pasien demam berdarah di Puskesmas Ngemplak Simongan. Gambar 8 menunjukkan analisis hasil pengujian data pasien demam berdarah puskesmas Ngemplak Simongan pada *rapidminer* menggunakan algoritma C4.5 memiliki akurasi yaitu 94.44 %.

Hasil pembentukan *decision tree* dalam bentuk description ditunjukkan oleh Gambar 9.

```

Tree
TROMBOSIT = RESIKO RENDAH: TIDAK (TIDAK=24, YA=0)
TROMBOSIT = RESIKO SEDANG
| SUHU TUBUH = SEDANG
| | BINTIK MERAH > 1: YA (TIDAK=0, YA=2)
| | BINTIK MERAH ≤ 1
| | | MIMISAN = TIDAK: TIDAK (TIDAK=3, YA=0)
| | | MIMISAN = YA: YA (TIDAK=1, YA=2)
| | SUHU TUBUH = TINGGI: YA (TIDAK=0, YA=29)
TROMBOSIT = RESIKO TINGGI: YA (TIDAK=0, YA=40)
    
```

**Gambar 9. Deskripsi Tree**

Gambar 9 di atas menunjukkan diskripsi dalam bentuk kalimat dari pohon keputusan dan menunjukkan hasil akhir dari penggunaan algoritma C4.5.

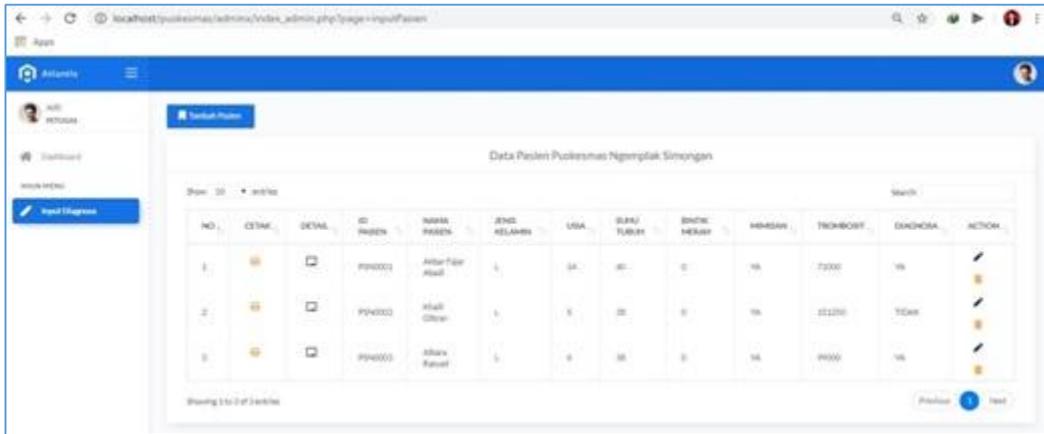
**4.8 Hasil**

Setelah melakukan penghitungan untuk mendapatkan nilai-nilai *entropy*, *gain*, *splitinfo*, dan *gainratio*, kemudian hasil penghitungan tersebut diaplikasikan menggunakan bahasa pemrograman berbasis web yaitu PHP (*Hypertext Preprocessor*). Berikut tampilan halaman utama Sistem Informasi Prediksi Penyakit Demam Berdarah dapat ditunjukkan pada Gambar 10.



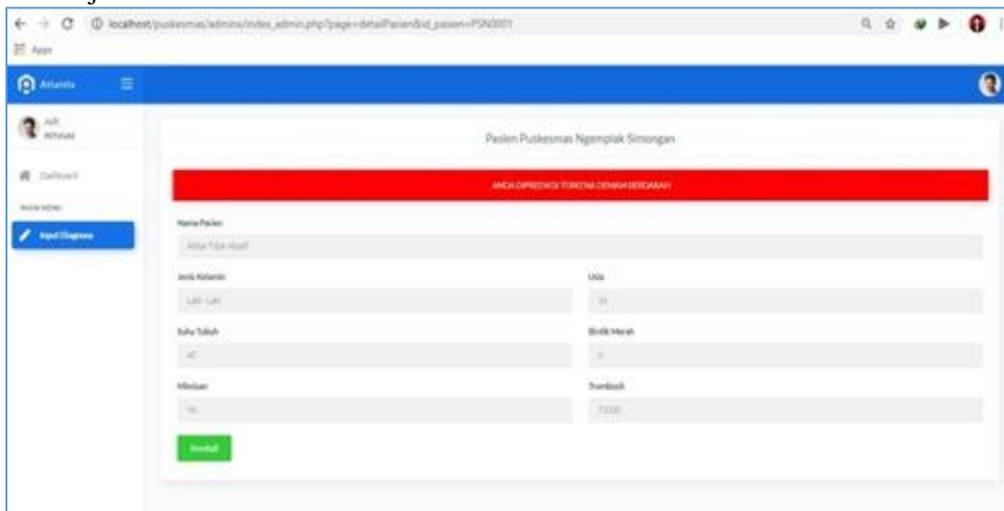
**Gambar 10. Tampilan Halaman Utama Sistem Informasi Prediksi Penyakit DBD**

Tampilan halaman utama program di atas terdapat pilihan untuk melakukan *input diagnosa*, disini petugas dapat mengelola sistem dengan menambah atau menginput data pasien baru. Halaman kelola data pasien ditunjukkan oleh Gambar 11.

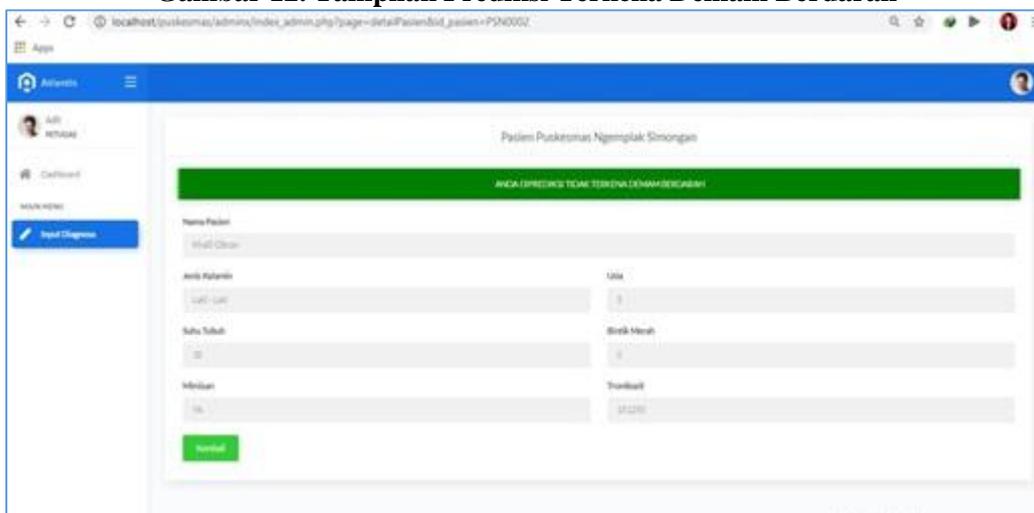


**Gambar 11. Tampilan Halaman Kelola Data Pasien**

Pada Gambar 11 di atas menampilkan data pasien yang telah diinputkan oleh petugas, yang kemudian data akan langsung tersimpan otomatis ke dalam database. Terdapat pilihan untuk melakukan tambah, edit, hapus, detail dan cetak data pasien. Tampilan prediksi pasien demam berdarah ditunjukkan oleh Gambar 12 dan Gambar 13.



**Gambar 12. Tampilan Prediksi Terkena Demam Berdarah**



**Gambar 13. Tampilan Prediksi Tidak Terkena Demam Berdarah**

Beberapa kode program untuk implementasi untuk implementasi algoritma C4.5 pada aplikasi di atas adalah sebagai berikut :

```
// ----- Cek Suhu Tubuh -----
if ($suhu_tubuh <= '38') {
    // Cek Bintik Merah
    if ($bintik_merah == 'YA') {
        // Cek Mimisan
        if ($mimisan == 'YA') {
            // ----- Cek trombosit -----

            // Jika trombositnya Tinggi
            if ($trombosit >= 45000 && $trombosit <= 90000) {
                $diagnosa = 'TIDAK';
            }
            // Jika trombositnya Sedang
            elseif ($trombosit >= 90000 && $trombosit <= 150000) {
                $diagnosa = 'TIDAK';
            }
            // Jika trombositnya Normal
            elseif ($trombosit >= 150000 && $trombosit <= 450000) {
                $diagnosa = 'TIDAK';
            } else {$diagnosa = 'ERROR'; }
        }
    }
    // Jika Tidak Mimisan
    else {
        // ----- Cek trombosit -----

        // Jika trombositnya Tinggi
        if ($trombosit >= 45000 && $trombosit <= 90000) {
            $diagnosa = 'TIDAK';
        }
        // Jika trombositnya Sedang
        elseif ($trombosit >= 90000 && $trombosit <= 150000) {
            $diagnosa = 'TIDAK';
        }
        // Jika trombositnya Normal
        elseif ($trombosit >= 150000 && $trombosit <= 450000) {
            $diagnosa = 'TIDAK';
        } else {$diagnosa = 'ERROR'; }
    }
}
```

## 5 KESIMPULAN

Berdasarkan pembahasan yang telah dijelaskan, maka dapat ditarik kesimpulan bahwa implementasi model berupa aplikasi pengujian adalah prediksi penyakit demam berdarah. Setelah petugas menginputkan nilai suhu tubuh, bintik merah, mimisan dan trombosit maka akan diperoleh nilai akurasi sebesar 94.44% sehingga aplikasi program yang dibangun dapat digunakan secara benar dan dapat dipertanggungjawabkan. Masukan dan pertimbangan pada aplikasi ini dapat lebih berintegrasi dengan sistem rekam medis dengan Rumah Sakit (rujukan) dan dapat dikembangkan dengan algoritma lain dalam data mining seperti Naïve Bayes, KNN dan lain-lain.

## REFERENSI

- [1] S. Sundaramurthy and P. Jayavel, "A hybrid Grey Wolf Optimization and Particle Swarm Optimization with C4.5 approach for prediction of Rheumatoid Arthritis," *Appl. Soft Comput. J.*, vol. 94, pp. 1–11, 2020.
- [2] K. R. Pradeep and N. C. Naveen, "Lung Cancer Survivability Prediction based on Performance Using Classification Techniques of Support Vector Machines, C4.5 and Naive Bayes Algorithms for Healthcare Analytics," *Procedia Comput. Sci.*, vol. 132, pp. 412–420, 2018.
- [3] A. Joshuva, R. S. Kumar, S. Sivakumar, G. Deenadayalan, and R. Vishnuvardhan, "An insight on VMD for diagnosing wind turbine blade faults using C4 . 5 as feature selection and discriminating through multilayer perceptron," *Alexandria Eng. J.*, pp. 1–17, 2020.
- [4] R. Benkercha and S. Moulahoum, "Fault detection and diagnosis based on C4.5 decision tree algorithm for grid connected PV system," *Sol. Energy*, vol. 173, no. April, pp. 610–634, 2018.

Saifur Rohman Cholil dkk : Prediksi Penyakit Demam Berdarah Di Puskesmas Ngemplak Simongan Menggunakan Algoritma C4.5

- [5] L. Han, W. Li, and Z. Su, "An assertive reasoning method for emergency response management based on knowledge elements C4.5 decision tree," *Expert Syst. Appl.*, vol. 122, pp. 65–74, 2019.
- [6] P. Kudła and T. P. Pawlak, "One-class synthesis of constraints for Mixed-Integer Linear Programming with C4.5 decision trees," *Appl. Soft Comput. J.*, vol. 68, pp. 1–12, 2018.
- [7] I. D. Mienye, Y. Sun, and Z. Wang, "Prediction performance of improved decision tree-based algorithms: A review," *Procedia Manuf.*, vol. 35, pp. 698–703, 2019.
- [8] X. Meng, P. Zhang, Y. Xu, and H. Xie, "Construction of decision tree based on C4.5 algorithm for online voltage stability assessment," *Int. J. Electr. Power Energy Syst.*, vol. 118, no. December 2019, pp. 1–8, 2020.
- [9] E. Ermawati, "Algoritma Klasifikasi C4.5 Berbasis Particle Swarm Optimization Untuk Prediksi Penerima Bantuan Pangan Non Tunai," vol. 8, no. September, pp. 513–528, 2019.
- [10] N. Azwanti and E. Elisa, "Analisis Pola Penyakit Hipertensi Menggunakan Algoritma C4 . 5," vol. 2, 2019.
- [11] J. Eska, "Penerapan Data Mining Untuk Prekdiksi Penjualan Wallpaper Menggunakan Algoritma C4.5 STMIK Royal Ksiaran," *JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 2, pp. 9–13, 2016.
- [12] S. Haryati, A. Sudarsono, and E. (2015) Suryana, "Implementasi Data Mining untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5," *J. Media Infotama*, vol. 11, no. 2, pp. 130–138, 2015.
- [13] R. Nofitri and N. Irawati, "Analisis Data Hasil Keuntungan Menggunakan Software Rapidminer," vol. V, no. 2, pp. 199–204, 2019.
- [14] G. G. Maulana, "Pembelajaran Dasar Algoritma Dan Pemrograman Menggunakan El-Goritma Berbasis Web," *J. Tek. Mesin*, vol. 6, no. 2, p. 8, 2017.
- [15] T. Tukino, "Penerapan Algoritma C4.5 Untuk Memprediksi Keuntungan Pada PT SMOE Indonesia," *J. Sist. Inf. Bisnis*, vol. 9, no. 1, p. 39, 2019.
- [16] F. F. Harryanto and S. Hansun, "Penerapan Algoritma C4.5 untuk Memprediksi Penerimaan Calon Pegawai Baru di PT WISE," *J. Tek. Inform. Dan Sist. Inf.*, vol. 3, no. 2, pp. 95–103, 2017.
- [17] A. Syafnur, "Analisis Dengan Metode Klasifikasi Menggunakan Decission Tree Untuk Memprediksi Penentuan Resiko Kredit Bank," vol. IV, no. 1, 2017.