

# PENERAPAN ALGORITME J48 UNTUK PREDIKSI PENYAKIT DEMAM BERDARAH

Tri Astuti<sup>1</sup>, Suciati<sup>2</sup>, Inke Mujiati<sup>3</sup>, Dian Ayu<sup>4</sup>, Vivit Ristianah<sup>5</sup>, Widhi Asty Lestari<sup>6</sup>

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

<sup>23456</sup>Program Studi Sistem Informasi

STMIK AMIKOM Purwokerto

Jl. Let. Jend. Pol Sumarto Purwokerto Indonesia

triasiasti@amikompurwokerto.ac.id<sup>1</sup>, suciati.almahi@gmail.com<sup>2</sup>, inke.bebe@yahoo.com<sup>3</sup>,

diaaaanayu@gmail.com<sup>4</sup>, vristianah@yahoo.com<sup>5</sup>, widhiasty25@gmail.com<sup>6</sup>

## ABSTRAK

*Demam Berdarah Dengue (DBD) merupakan penyakit infeksi virus yang disebabkan oleh gigitan nyamuk Aedes aegypti betina. Penyakit demam berdarah masih termasuk dalam kategori penyakit yang berbahaya, bahkan dapat menyebabkan kematian. Lamanya proses diagnosis penyakit tersebut menjadi hambatan bagi seseorang untuk mendapatkan diagnosis akan penyakit yang diderita secara dini. Penelitian ini dilakukan guna memprediksi penyakit demam berdarah menggunakan metode algoritme J48 yang di implementasikan pada data pasien penyakit demam berdarah. Dari hasil analisis menggunakan cross validation vold 5 diperoleh nilai akurasi 75.8333%, dan vold 10 menghasilkan nilai akurasi sebesar 80%.*

*Kata Kunci: Decision tree, Demam Berdarah, J48*

## PENDAHULUAN

Penyakit demam berdarah di Indonesia merupakan penyakit yang mendorong angka kematian yang cukup tinggi, sehingga sampai sekarang penyakit tersebut ditakuti oleh manusia. Oleh karena itu penyakit demam berdarah perlu diprediksi yaitu dengan menggunakan klasifikasi data mining sehingga praktisi kesehatan dalam pengambilan keputusan bisa lebih tepat dan akurat. Salah satu algoritme yang dapat diterapkan dalam proses klasifikasi adalah decision tree J48.

Algoritme J48 dalam klasifikasi sebagai teknik dalam data mining yang digunakan dalam penelitian untuk mendeteksi seorang pasien terkena DBD dilihat dari data pasien pada suatu rumah sakit berdasarkan ciri-ciri dan gejala pasien. Dengan pengelompokkan ini akan mempermudah manajemen dan pengawasan penyakit untuk melihat apakah penyakit DBD berdasarkan ciri-ciri, gejala, dan kondisi tubuh pasien tersebut. Dalam penulisan penelitian ini penulis akan menggunakan algoritme J48 dikarenakan menurut penelitian sebelumnya telah

disebutkan bahwa teknik decision tree lebih fleksibel dan membuat metode ini atraktif, khususnya karena memberikan keuntungan berupa visualisasi saran (dalam bentuk decision tree) sehingga prosedur klasifikasinya dapat diamati (Gorunescu, 2011).

Penelitian yang dilakukan dalam menghasilkan keputusan dan prediksi penyakit demam berdarah berdasarkan 120 kasus dan 16 Atribut dengan decision tree mempunyai kelebihan-kelebihan dalam mengolah dataset penyakit demam berdarah. Berdasarkan atas penelitian diatas, peneliti akan memilih metode decision tree dalam memprediksi penyakit demam berdarah. Dalam penelitian ini akan dilakukan penerapan algoritme decision tree (J48) menggunakan metode data mining dengan mengoptimal atribut-atribut yang berasal dari dataset yang terpercaya untuk memprediksi penyakit demam berdarah dengan tujuan agar akurasi menjadi meningkat. Industri kesehatan memiliki sejumlah besar data kesehatan, namun sebagian besar data tersebut tidak diolah untuk mengetahui informasi tersembunyi untuk dijadikan pengambilan keputusan yang efektif oleh para praktisi kesehatan. Pengambilan keputusan atas dasar data dan informasi yang akurat akan menghasilkan keputusan dan prediksi penyakit menjadi tepat sasaran.

## **TINJAUAN PUSTAKA**

### a. Data Mining

Data mining adalah proses pengolahan informasi dari sebuah database yang besar, meliputi proses ekstraksi, pengenalan, komprehensif, dan penyajian informasi sehingga dapat digunakan dalam pengambilan keputusan.

### b. Decision tree

Decision tree adalah salah satu metode klasifikasi yang paling populer karena mudah untuk diinterpretasi oleh manusia. decision tree adalah model prediksi menggunakan struktur pohon atau struktur berhirarki. Konsep dari decision tree adalah mengubah data menjadi pohon keputusan dan aturan-aturan keputusan. Manfaat utama dari penggunaan decision tree adalah kemampuannya untuk mem-break down proses pengambilan keputusan yang kompleks menjadi lebih simpel sehingga pengambil keputusan akan lebih menginterpretasikan solusi dari suatu permasalahan.

Salah satu algoritme induksi pohon keputusan yaitu ID3 (Iterative Dichotomiser 3). ID3 dikembangkan oleh J. Ross Quinlan. Dalam prosedur algoritme ID3, input berupa sampel training, label training dan atribut. Algoritme C4.5 merupakan pengembangan dari ID3. Sedangkan pada perangkat lunak open source WEKA mempunyai versi sendiri C4.5 yang dikenal sebagai J48.

Pohon dibangun dengan cara membagi data secara rekursif hingga tiap bagian terdiri dari data yang berasal dari kelas yang sama. Bentuk pemecahan (split) yang digunakan untuk membagi data tergantung dari jenis atribut yang digunakan dalam split. Algoritme C4.5 dapat menangani data numerik (kontinyu) dan diskret. Split untuk atribut numerik yaitu mengurutkan contoh berdasarkan atribut kontinyu A, kemudian membentuk minimum permulaan (threshold) M dari contoh-contoh yang ada dari kelas mayoritas pada setiap partisi yang bersebelahan, lalu menggabungkan partisi-partisi yang bersebelahan tersebut dengan kelas mayoritas yang sama. Split untuk atribut diskret A mempunyai bentuk value (A)  $\in X$  dimana  $X \subset \text{domain}(A)$ .

Jika suatu set data mempunyai beberapa pengamatan dengan missing value yaitu record dengan beberapa nilai variabel tidak ada, Jika jumlah pengamatan terbatas maka atribut dengan missing value dapat diganti dengan nilai rata-rata dari variabel yang bersangkutan. [Santosa,2007]

## **METODE PENELITIAN**

Metode yang digunakan adalah metode algoritme J48 digunakan untuk mendeteksi penyakit demam berdarah pada pasien opname di Rumah Sakit Umum Wijaya Kusuma Daerah Kabupaten Banyumas. Sesuai dengan sumber data dalam penyusunan peneltian ini maka dalam penulisan dan pengumpulan data, penulis menggunakan beberapa metode, antara lain:

### **a. Pengumpulan Data**

Pengumpulan dilakukan melalui wawancara dan juga observasi lapangan. Adapun data yang diperoleh adalah data berupa data kasus DBD. Selain itu juga data-data yang terkait dengan hal atau faktor-faktor yang mempengaruhi daerah terkena DBD seperti data usia, jenis kelamin, demam, torniket, pendarahan, hepatomegaly, syok, hipotensi, nyeri, mual.

b. Analisis

Data yang diperoleh dilakukan teknik analisis yaitu data akan dipilih dan dilakukan pre-prosesing data. Adapun tahapan data pre-prosesing yang dilakukan adalah sebagai berikut :

1. Pembersihan data (cleaning data), yaitu melakukan pembersihan terhadap data yang telah diperoleh dengan cara menghilangkan data yang dianggap mempunyai nilai yang sama, data yang tidak lengkap, dan juga data yang tidak sesuai.
2. Setelah data diperoleh maka data data tersebut akan dilakukan intregasi dan transformasi data, yang dilakukan pada tahapan ini adalah data akan diintegrasikan dengan data yang sesuai dan ditransformasikan kedalam satu kesatuan hingga menjadi data yang saling terintegrasi dan berhubungan antara data yang satu dengan yang lainya. Adapun hasil data terintegrasi dapat dilihat pada Tabel 1.

Tabel 1. Daftar *attribute* demam berdarah *dataset*

No.	Attribute	Nilai Attribute
1.	Usia	Years
2.	Jenkel	{1, 0} laki-laki, perempuan
3.	demam	{1, 0} ya, tidak
4.	torniket	{1, 0} ya, tidak
5.	pendarahan	{1, 0} ya, tidak
6.	hepatomegali	{1, 0} ya, tidak
7.	Syok	{1, 0} ya, tidak
8.	hipotensi	{1, 0} ya, tidak
9.	Nyeri	{1, 0} ya, tidak
10.	Mual	{1, 0} ya, tidak
11.	epigastrium	{1, 0} ya, tidak
12.	Lemas	{1, 0} ya, tidak

13.	trombosit	Real
14.	hemoglobin	Real
15.	hematokrit	Real
16.	Class	{0, 1, 2, 3} Sehat, Demam Dengue, Dengue, Syndrome Dengue.

c. Tahap Penyusunan *decision tree*

*Decision tree* J48 merupakan implementasi dari algoritme C4.5 yang memproduksi *decision tree*. Ini merupakan standar algoritme yang digunakan dalam *machine learning*. *Decision tree* merupakan salah satu algoritme klasifikasi dalam data mining. Algoritme klasifikasi merupakan algoritme yang secara induktif dalam pembelajaran dalam mengkonstruksikan sebuah model dari dataset yang belum diklasifikasikan (pre classified dataset). Setiap data dari item berdasarkan dari nilai dari setiap atribut. Klasifikasi dapat dilihat sebagai mapping dari sekelompok set dari atribut dari kelas tertentu. *Decision tree* mengklasifikasikan data yang diberikan menggunakan nilai dari atribut.

## HASIL DAN PEMBAHASAN

1. Penelitian ini menggunakan dataset demam berdarah rumah sakit Wijaya Kusuma untuk proses imputasi. Dataset ini terdiri dari 120 kasus dengan 15 kondisi atribut dan satu atribut keputusan. Daftar atribut pada dataset demam berdarah dapat dilihat pada Tabel I. Preprocessing dilakukan dengan membuat dataset lengkap merupakan proses pertama kali dilakukan dalam imputasi.

Tabel 2. Data yang belum dihilangkan atributnya:

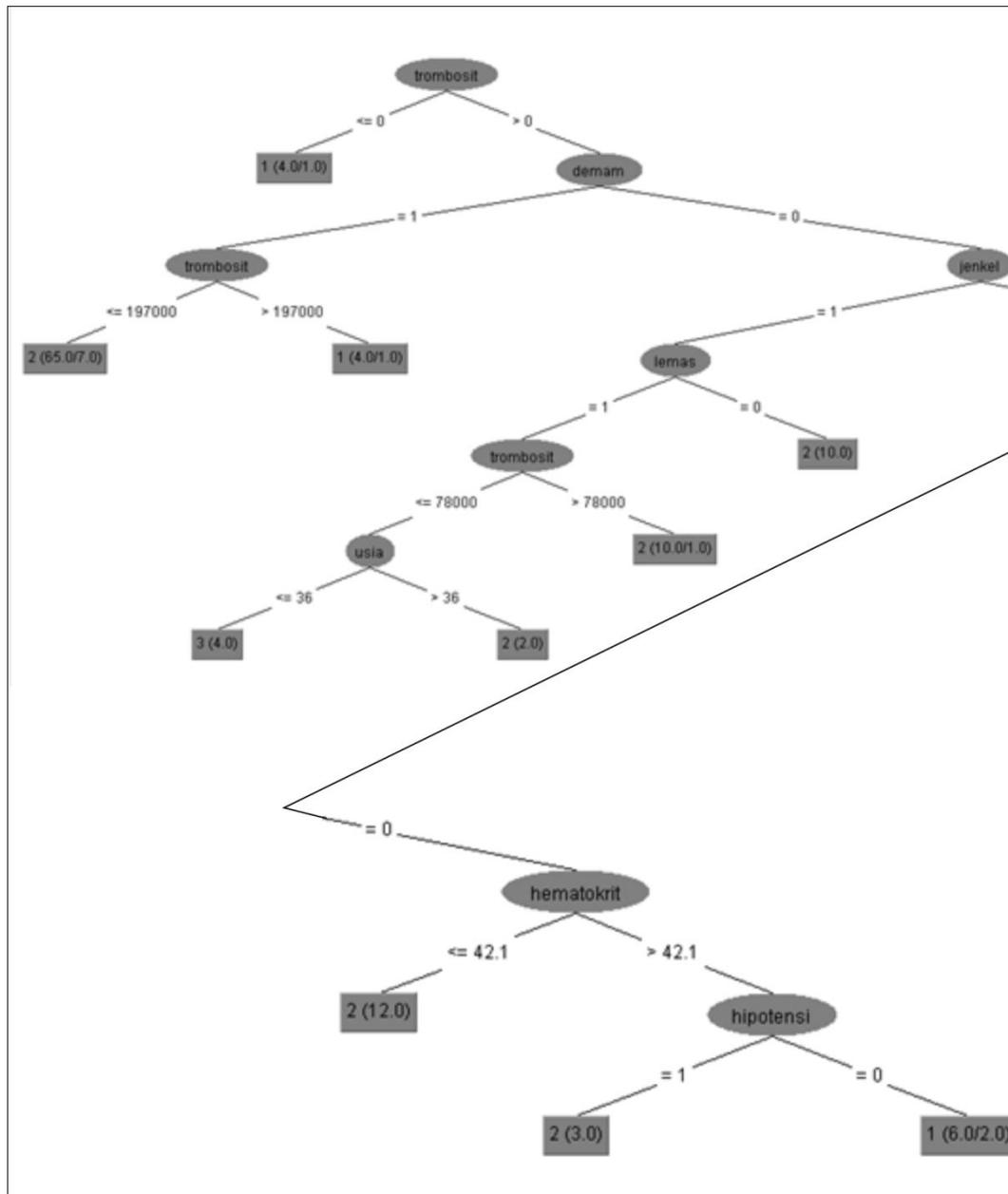
<i>Attribute</i>	<i>Nilai Attribute</i>
Usia	Years
Jenkel	{1, 0} laki-laki, perempuan
Demam	{1, 0} ya, tidak
Torniket	{1, 0} ya, tidak
Pendarahan	{1, 0} ya, tidak
Hepatomegali	{1, 0} ya, tidak
Syok	{1, 0} ya, tidak
Hipotensi	{1, 0} ya, tidak
Nyeri	{1, 0} ya, tidak
Mual	{1, 0} ya, tidak
Epigastrium	{1, 0} ya, tidak
Lemas	{1, 0} ya, tidak
Trombosit	Real
Hemoglobin	Real
Hematokrit	Real
Class	{0, 1, 2, 3} Sehat, Demam Dengue, Dengue, Syndrome Dengue.

2. Untuk mendapatkan pohon keputusan maka kita perlu menghilangkan beberapa attribute.

Tabel 3. Data yang sudah dihilangkan atributnya

<i>Attribute</i>	<i>Nilai Attribute</i>
Usia	Years
Jenkel	{1, 0} laki-laki, perempuan
Demam	{1, 0} ya, tidak
Hipotensi	{1, 0} ya, tidak
Mua	{1, 0} ya, tidak
epigastrium	{1, 0} ya, tidak
Lemas	{1, 0} ya, tidak
trombosit	Real
hemoglobin	Real
hematokrit	Real
Class	{0, 1, 2, 3} Sehat, Demam Dengue, Dengue, Syndrome Dengue.

3. Nilai akurasi yang didapatkan dari penelitian ini dengan menggunakan cross validation volt 5 sebesar 75.8333% dan volt 10 nilai akurasinya 80%.
4. Informasi nilai yang terletak dalam kurung pada setiap akhir node menunjukkan banyaknya data yang dilatih dan jumlah data yang salah dikenali. Apabila hanya terdapat satu nilai saja menunjukkan bahwa semua data di akhir node tersebut dapat diklasifikasikan dengan baik semuanya. Dari 120 data, didapatkan nilai akurasinya sebesar 80%. Setelah menghilangkan beberapa attribute maka menghasilkan pohon keputusan:



Gambar 1. Decision tree J48 pada dataset DBD

## **KESIMPULAN DAN SARAN**

### a. Kesimpulan

Penelitian ini dilakukan menggunakan algoritme J48 dari hasil eksperimen menunjukkan tingkat akurasi sebesar 75,833% menggunakan cross validation vold 5 dan hasil 80% menggunakan vold 10 dalam mendeteksi klasifikasi penyakit demam berdarah. Penelitian dilakukan dari tahap awal hingga pengujian menggunakan penerapan algoritme untuk memberikan jawaban atas permasalahan sulitnya mendeteksi klasifikasi penyakit demam berdarah pada pasien opname. Berdasarkan hasil penelitian, menyimpulkan bahwa, algoritme mudah dalam mendeteksi penyakit demam berdarah pada pasien opname.

### b. Saran

Berdasarkan hasil penelitian, penerapan algoritme dalam mendeteksi penyakit demam berdarah pada pasien opname ini dapat membawa efek positif dalam proses mendeteksi penyakit demam berdarah, namun terdapat beberapa hal yang perlu penulis sarankan antara lain:

1. Dalam pengukuran akurasi data dapat lebih untuk ditingkatkan.
2. Ada baiknya apabila dilakukan pengujian data dengan teknik algoritme menggunakan lebih dari satu metode algoritme untuk mendapatkan nilai akurasi yang lebih tinggi.

## **DAFTAR PUSTAKA**

- Cetiner B. Gultekin, Sari Murat dan M.ABuras Hani, 2009, “ Recognition of Dengue Diseases Patterns Using Articial Neural Network,” Department of Manufacturing and Materials Engineering,Kualumpur, Malaysia.
- Iancu Eugen, Iancu Ionela, Sfredel Veronica, 2010, “Predictive Control of Blood Glucose in Diabetes Mellitus Patients” Department of Automation, University of Craiova.
- Rahim H.Abdul., Ibrahim., Taib M.N., 2010, “System Identification of Non Linear Autoregressive Models in Monitoring Dengue Infection,” Universiti Teknologi Malaysia.

Santoso Budi, "Buku Data Mining: Data Mining, Teknik Pemanfaatan Data untuk Keperluan Bisnis," Graha.

Sapna, S, Ms., Tamilarasi, A., 2012, "Fuzzy Relational Equation in Preventing Diabetic Heart Attack", Dept. of Master of Computer Applications, India, 2009. Jurnal Teknologi Informasi, Volume 8 Nomor 1, April 2012, ISSN 1414-999966 <http://research.pps.dinus.ac.id> email redaksi: stefanus@pps.dinus.ac.id

Tan Philip, Cing Tay Joc, 2006, "Evolving Boolean Network to Find Intervention Points in Dengue Pathogenesis" Nanyang Technological University, Singapore.