

**IMPLEMENTASI ALGORITMA C4.5
UNTUK MENGLASIFIKASI PENYAKIT TIPES DAN DBD
(STUDI KASUS RUMAH SAKIT UMUM RAA
SUWONDO PATI)**

Preparation of Papers for IEEE JOURNAL

First A. Author, Second B. Author, Jr., and Third C. Author, *Member, IEEE*

Abstract Dengue Fever is one of the major health problems in some countries, including Indonesia. Indonesia is included in category 'A' in the DBD stratification by WHO in 2001 which indicated high rates of treatment in hospital and deaths due to dengue. That's because late in the handling or it could be a doctor one because diagnosing dengue disease has symptoms similar to other diseases, especially typhoid.

The purpose of this study was to classify the patient's disease using the most influential variable for diseases typhoid and dengue. The object of this study is the laboratory test results, patient characteristics, symptoms, and physical conditions in medical pasien obtained from National Starch Hospital Partners as many as 50 data of patients, 22 patients and 28 patients of dengue fever typhoid.

Variables used as hematokrit of laboratory tests and test widal test, but it is also used variable fever, longer fever, tongue color, and platelets.

Formation of dengue and typhoid disease classification using RapidMiner applications developed since 2001 by Ralf Klöckner, Ingo Meirswa, and Simon Fischer in Artificial Intelligence Unit of the Technical University of Dortmund. The algorithm that is used is the C4.5 Decision Tree models are represented in the form of rule, of a model formed the most influential variable is hematokrit.

Keywords: Decision Tree, C4.5 algorithms, Dengue, Types

I. PENDAHULUAN

(Kehadiran data mining dilatar belakangi oleh berlimpahnya data (overload data) yang dialami oleh berbagai institusi, perusahaan atau organisasi. Berlimpahnya data ini merupakan akumulasi data transaksi yang terekam bertahun-tahun.

Klasifikasi merupakan salah satu fungsi yang umum diterapkan dalam data mining. Masalah klasifikasi saat ini banyak sekali ditemukan dalam kehidupan sehari-hari misalnya dalam bidang pemasaran, untuk memprediksi penjualan dan jadwal distribusi diantara banyak outlet, dalam bidang broadcasting, untuk memprediksi program terbaik apa yang ditayangkan selama waktu tayang utama, dan terakhir sering dipakai yaitu dalam bidang kedokteran, misalnya untuk mendiagnosis penyakit.

Banyak orang yang salah dalam mendiagnosis penyakit types dan DBD, karena keduanya memiliki ciri-ciri dan gejala yang hampir sama sehingga sulit membedakannya, namun sebenarnya berbeda. Pada penelitian ini, dengan menggunakan data pasien rawat inap suatu rumah sakit untuk penyakit demam berdarah dan types dapat dibuat suatu bentuk pemodelan agar kita dapat

melihat dan mengetahui faktor apa yang sangat berpengaruh dalam mendiagnosis penyakit tersebut. Problem seperti ini dalam data mining termasuk dalam jenis klasifikasi.

Pada skripsi ini diusulkan implementasi algoritma C4.5 dalam klasifikasi sebagai teknik dalam data mining yang digunakan dalam penelitian untuk mengklasifikasi seorang pasien masuk dalam kategori types atau 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 tersebut masuk dalam kategori types atau DBD berdasarkan ciri-ciri, gejala, dan kondisi tubuh pasien tersebut.

Perangkat lunak RapidMiner digunakan sebagai alat bantu dalam proses klasifikasi data. Dalam penulisan skripsi ini penulis akan menggunakan algoritma C4.5 dikarenakan menurut penelitian sebelumnya telah disebutkan bahwa teknik decision tree lebih kefleksibel dan membuat metode ini atraktif, khususnya karena memberikan keuntungan berupa fisualisasi saran (dalam bentuk decision tree) yang terbuat sehingga prosedur klasifikasinya dapat diamati (Gorunescu, 2011).

conference, please observe the conference page limits.

II. METODE YANG DIUSULKAN

Sesuai dengan sumber data dalam penyusunan tugas akhir ini maka dalam penulisan dan pengumpulan data, penulis menggunakan beberapa metode, antara lain :

2.1 Studi Pustaka

Metode ini menggunakan teknik pengumpulan data melalui file-file yang berkaitan dengan penyakit Demam Berdarah Dengue dan Tipes, file-file tersebut didapat melalui buku-buku dan beberapa jurnal yang membahas tentang klasifikasi Demam Berdarah Dengue dan Tipes. Semua itu dapat digunakan sebagai data pendukung dalam penulisan tugas akhir ini.

2.2 Wawancara

Metode ini dilakukan untuk menemukan sebuah informasi dengan cara *face to face* dengan narasumber langsung, dan memberikan pertanyaan seputar topik yang akan di bahas yaitu mengenai diagnosa penyakit Demam Berdarah Dengue dan Tipes. Wawancara dilakukan bersama salah satu dokter dari Rumah Sakit Umum Daerah RAA Suwondo Pati yaitu dr. Albert Tri Rustamaji, Sp.PD.

2.2.1 Penelitian Dan Mengunjungi Situs

Penelitian adalah teknik yang sering digunakan berdasarkan studi terhadap sistem serupa. Kunjungan situs merupakan bentuk penelitian yang khusus, dengan menjelajahi internet dapat memperoleh informasi yang tak terduga. Dalam hal ini penulis mencari beberapa teori yang dibutuhkan dalam menyelesaikan tugas akhir melalui media internet. Situs – situs yang dikunjungi adalah yang berkaitan dengan penyakit Demam Berdarah Dengue dan Tipes.

2.3 Metode Analisis Data

Adapun untuk menganalisis data dalam penerapan data mining ini menggunakan proses tahapan knowledge discovery in databases (KDD) yang terdiri dari Database, Data Cleaning, Data integration, Task relevan data, Data transformation, Data mining, Pattern evolution, knowledge (Han J & Kamber. 2006) :

1. Database

Koleksi data yang saling berhubungan untuk dipergunakan secara bersama kemudian dirancang untuk memenuhi kebutuhan informasi.

2. Data Cleaning

Pada umumnya, data yang diperoleh, baik dari database suatu rumah sakit maupun eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau hanya sekedar salah ketik. Data-data yang tidak relevan itu juga lebih baik dibuang karena keberadaannya bisa mengurangi mutu atau akurasi dari hasil data mining nantinya. Pembersihan data juga akan mempengaruhi performansi dari system data mining karena data yang akan ditangani akan berkurang jumlah dan kompleksitasnya.

3. Data Integration

Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut usia, suhu tubuh, tes widal hematokrit dan lain lain. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil data yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya.

4. Task Relevan Data

Setelah semua sumber data digabung atau diintegrasikan menjadi satu keseluruhan database, maka tahap selanjutnya adalah melakukan task relevant data. Pada tahap ini melakukan relevansi atribut dari data yang relevant atau yang sesuai dengan target atau output yang akan dihasilkan.

5. Data Transformation

Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses coding dalam KDD merupakan proses kreatif dan sangat

tergantung pada jenis atau pola informasi yang akan dicari dalam database.

6. Data Mining

Data mining adalah proses mengeksplorasi dan menganalisa data dalam jumlah yang besar yang bertujuan untuk menemukan suatu pola atau informasi yang menarik dari data yang tersimpan dalam jumlah yang besar dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD

secara keseluruhan. Tahap ini merupakan inti dari tahapan KDD yang dilakukan untuk menganalisis data yang telah dibersihkan.

7. Pattern Evaluation

Dalam tahap ini, merupakan hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai dengan hipotesa, ada beberapa alternatif yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki data mining lain yang lebih sesuai, atau menerima hasilnya sebagai suatu hasil yang diluar dugaan yang mungkin bermanfaat.

8. Knowledge

Tahap terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisa yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak paham mengenai data mining. Karenanya presentasi hasil data mining dalam bentuk pengetahuan yang bisa dipahami semua orang adalah itu tahapan yang diperlukan dalam proses data mining. Dalam presentasi ini, visualisasi juga bisa membantu mengkomunikasikan hasil data mining.

III. IMPLEMENTASI

3.1 Implementasi Dalam Aplikasi

a. Tampilan awal

Dalam membuat tampilan awal aplikasi ini menggunakan bahasa pemrograman PHP dan di simpan dengan nama *index.php* dan headernya dengan nama *kepala.php*.

Dari source code dihasilkan tampilan awal seperti berikut :



Gambar 4.17 Tampilan Awal

b. Tampilan pertanyaan

Tampilan pertanyaan disimpan dengan nama *diagnosa.php*

Dari source code tersebut diperoleh hasil tampilan sebagai berikut :



Gambar 4.18 Tampilan Pertanyaan

Gambar tersebut akan muncul ketika pengguna memulai diagnosa. Tampilan diatas akan muncul dan sama pada setiap pertanyaan yang diberikan. Ketika pengguna menekan tombol lanjut maka akan lanjut ke pertanyaan berikutnya.

Pengguna tidak bisa menyelesaikan diagnosa sebelum menjawab semua pertanyaan yang diberikan.

c. Hasil diagnosa

Hasil diagnosa di simpan dalam file *solving.php*

Dari source code tersebut dihasilkan tampilan sebagai berikut :



Gambar 4.19 Tampilan Hasil Diagnosa

Pengguna akan kembali ke pertanyaan awal ketika menekan tombol “Kembali Melakukan Diagnosa”

d. Database

Source code diatas tidak akan berjalan jika tidak ada database, maka sebelum menjalankan aplikasi diatas dibuat data base dengan nama dbdtipes.sql dan di beri tabel sebagai berikut :

Gambar 4.20 Tabel Database Database yang berisi tabel di atas kemudian di beri alur rule yang ada. Dibawah ini adalah logika rule yang telah di masukan.

Gambar 4.21 Tabel Database dengan Rule

Dengan source code dan database yang telah dibuat maka program dapat dijalankan.

IV. HASIL & PEMBAHASAN

4.1 Hasil Pengujian Metode Decicion Tree C4.5

a. Confusion Matrix

Berdasarkan data training sebanyak 62 example set dengan 9 atribut berupa usia, suhu tubuh, Salmonella Typhi O, Salmonella Typhi H, dan Hematrokit yang dimodelkan dengan algoritma decision tree C4.5 diperoleh hasil akurasi sebanyak 88.71 % dengan rincian sebagai berikut

accuracy: 88.57% +/- 10.58% (mikroc: 88.71%)			
	true DBD	true Tipes	class precision
pred. DBD	37	4	90.24%
pred. Tipes	3	18	85.71%
class recall	92.50%	91.82%	

Gambar 4.22 nilai akurasi model C4.5

Jumlah *true positive (tp)* adalah 37 jumlah pasien yang di prediksi sakit DBD dan benar pasien tersebut menderita sakit DBD. Jumlah *false positive (fp)* adalah 4 pasien yang seharusnya sakit tipes namun diprediksi sakit DBD. Selain itu jumlah *true negative (tn)* adalah 18 jumlah pasien yang di prediksi dengan benar mengalami sakit tipes dan jumlah *false negative (fn)* adalah 3 pasien yang seharusnya sakit DBD namun di prediksi sakit tipes.

Dari tabel tersebut dapat diketahui bahwa tidak semua sakit pasien DBD dan tipes diprediksi dengan benar, oleh karena itu akurasi yang di hasilkan tidak dapat mencapai 100 %. Namun demikian akurasi yang di hasilkan masih cukup besar dan dapat dijadikan acuan untuk mendeteksi penyakit DBD dan tipes.

Berdasarkan hasil *confision matrix*, menunjukkan bahwa hasil akurasi adalah sebesar 88.71 %. Perhitungan dari *accuracy, sensitivity, specificity, PPV, dan NPV* adalah sebagai berikut :

$$\begin{aligned}
 \text{Acc} &= \frac{tp + tn}{tp + fp + tn + fn} = \frac{37 + 18}{37 + 4 + 18 + 3} = 0,8871 \\
 \text{Sensitivity} &= \frac{tp}{tp + fn} = \frac{37}{37 + 3} = 0,9250 \\
 \text{Specificity} &= \frac{tn}{tn + fp} = \frac{18}{18 + 4} = 0,8182 \\
 \text{PPV} &= \frac{tp}{tp + fp} = \frac{37}{37 + 4} = 0,9024 \\
 \text{NPV} &= \frac{tn}{tn + fn} = \frac{18}{18 + 3} = 0,8571
 \end{aligned}$$

Dari perhitungan diatas dapat diketahui bahwa nilai accuracy adalah 88,71% , nilai sensitivity adalah 92,50% , nilai specificity adalah sebesar 81,82% , nilai PPV adalah 90,24% , dan nilai NPV adalah 85,71%.

b. Pengujian dengan Aplikasi

Pengujian dengan aplikasi dilakukan dengan menambahkan 7 data baru yaitu :

Tabel 4.3 Dataset Baru

Jenis kelamin	Suhu tubuh	Salmonella Thypi O	Hematokrit	Hemoglobin	Usia	Penyakit
L	30	Negative	20	15	20	DBD
L	35	Negative	43	14	40	DBD
P	40	Negative	33	12	35	DBD
L	27	Positive	40	12	45	TIPES
P	24	Positive	41	12	20	TIPES
L	30	Positive	22	17	34	DBD
P	35	Negative	41	14	40	TIPES

Dari data baru di atas di masukan kedalam aplikasi dan hasilnya adalah sebagai berikut :

Tabel 4.4 Hasil pengujian dengan aplikasi

No.	Jenis kelamin	Suhu tubuh	Salmonella Thypi O	Hematokrit	Hemoglobin	Usia	Penyakit	Hasil
1.	L	30	Negative	20,2	15	20	DBD	✓
2.	L	35	Negative	43	14,2	40	DBD	✓
3.	P	40	Negative	33	12	35	DBD	✓
4.	L	27	Positive	40,3	12	45	TIPES	✓
5.	P	24	Positive	41	12,3	20	TIPES	✓
6.	L	30	Positive	22	17	34	DBD	✓
7.	P	35	Negative	41	14,3	40	TIPES	✓

Dari tabel diatas didapatkan hasil bahwa semua data baru yang ada dapat didiagnosa dengan benar oleh aplikasi yang di buat. Hal itu berarti aplikasi yang telah dibuat sesuai dengan rule yang didapat sebelumnya, dan aplikasi dapat digunakan oleh pengguna. Karena sifatnya yang online maka aplikasi dapat langsung digunakan oleh user ketika user telah melakukan tes laboratorium tanpa harus pergi kedokter. Namun demikian untuk mendapatkan penangan dan hasil yang lebih akurat dapat di tanyakan ke dokter yang menangani bidang tersebut.

V. PENUTUP

Kesimpulan yang didapatkan dari hasil penelitian ini adalah metode Decision Tree C4.5 dapat diimplementasikan untuk pengelompokan data penyakit DBD dan Tipes dengan menggunakan 6 parameter penyakit DBD dan Tipes yaitu jenis kelamin, usia, hemoglobin, hematokrit, tinggi demam dan *Salmonella Thypi O*. Berdasarkan hasil pengujian sistem hasil rata – rata *Confusion Matrix* pada jumlah 62 data diperoleh hasil *accuracy* adalah 88,71% dan berdasarkan hasil pengujian dengan menggunakan aplikasi diperoleh hasil bahwa semua

data baru yang di berikan dapat didiagnosa dengan benar dan tingkat accuracy mencapai 100%.

REFERENCES

- Fs Rodiyatul dan Bayu Adhi Tama. Universitas Sriwijaya : Implementasi Teknik Data Mining Di Dalam Analisis Penyakit Diabetes Mellitus Tipe II Menggunakan Decision Tree. 2014.
- Gorunescu, F.2011. Data Mining : Concepts, Model and Techniques. New York : Springer-Verlag.
- Han J dan Kamber. Data Mining: Concepts And Techniques Second Edition, 2006.
- Kurniawan , Edi. Institut Teknologi Sepuluh November Surabaya: Analisa Rekam Medis Untuk Menentukan Pola Kelompok Penyakit Menggunakan Klasifikasi Decision Tree J48. 2011.
- Prasetyo, E.2013. Data Mining – Konsep dan Aplikasi Menggunakan Matlab. Yogyakarta : Andi Offset.
- Quinlan, J.R. 1986. "Induction of Decision Trees". Machine Learning 1:81-106.
- Quinlan, J.R. 1996. "Improved Use of Continuous Attributes in C4.5". *Journal of Artificial Intelligence Research* 4 : 77-90.
- Romario, Sandro. Universitas Binadarma: Penerapan Data Mining Pada Bsup Dr. Moh Hoesin Sumatera Selatan Untuk Pengelompokan Hasil Diagnosa Pasien Pengguna Asuransi Kesehatan Miskin (Askin). 2013.
- Shynk, J.J.T.1990. "Performance Surfaces of A Single-Layer Perceptron". *IEEE Transaction on Neural Networks* 1 : 268-274. TIPES ✓
- Supriyadi, Didi. Universitas Diponegoro Semarang : Sistem Informasi Penyebaran Penyakit Demam Berdarah Menggunakan Metode Jaringan Syaraf Tiruan Backpropagation" 23 Agustus 2014.