

PREDIKSI PENYAKIT DEMAM BERDARAH DAN TYPHUS DENGAN ALGORITMA C5.0

Tri Haryanto¹, Shaufiah², Fhira Nhita³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Gejala penyakit demam berdarah (DBD) dan typhus memiliki beberapa kesamaan. Kedua penyakit ini ditandai dengan demam yang cukup tinggi sehingga mengganggu aktivitas penderitanya. Jika salah mengetahui jenis penyakit yang diderita bisa mengakibatkan kesalahan penanganan pada penderita, bahkan dapat menyebabkan kematian. Oleh karena itu penelitian ini bertujuan untuk melakukan klasifikasi terhadap kedua penyakit tersebut berdasarkan pada gejala dari penyakitnya yang di derita oleh pasien. Untuk melakukan klasifikasi terhadap kedua penyakit ini, dibuat suatu model pohon keputusan untuk memprediksi penyakit pasien dan melihat variabel yang paling berpengaruh atas penyakit pasien dengan kategori typhus dan DBD. Data penelitian ini adalah medical record rumah sakit yang berupa data hasil uji laboratorium pasien, ciri-ciri, gejala, dan kondisi fisik pasien yang di dapat dari medical record rumah sakit. Data yang digunakan adalah hasil laboratorium pasien yang diambil pada saat pemeriksaan pertama pasien datang sebelum dinyatakan memiliki penyakit typhus, DBD atau komplikasi kedua penyakit tersebut. Variabel bebas atau variabel input (prediktor) pada penelitian ini adalah pemeriksaan awal pasien yaitu meliputi kategori usia, demam, lama demam, trombosit, hematokrit, dan uji widal pasien. Sedangkan kondisi pasien yang akan diprediksi yaitu status penyakit pasien typhus atau DBD. Algoritma yang digunakan ialah C5.0 yang menghasilkan model yang direpresentasikan dalam bentuk aturan (rule). Dari model yang dihasilkan diperoleh akurasi diatas 84%.

Kata Kunci : Demam Berdarah, Typhus, Algoritma C5.0

Abstract

Symptoms of dengue fever or dengue and typhoid have some similarities. Both diseases are characterized by a high fever that disrupt the activity of sufferers. If mistakenly identify disease recognizes the type of illness can lead mistakenly in treating patients, it can even cause death. Therefore, this study aims to classify the two diseases based on the symptoms of the illness suffered by the patient. To perform the classification of the two diseases, is a decision tree model to predict the patient's illness and saw the most influential variable on disease patients with typhoid and dengue category. The data of this research is the laboratory test results of patients, characteristics, symptoms, and physical condition of patients in the can from the hospital medical record. The data used is laboratory results of patients were taken during examination the patient come first before it had typhoid or dengue disease. The independent variables or input variables (predictors) in this study was the examination of patients which include age categories, fever, duration of fever, platelets, hematokrit, and test widal patients. While the condition of patients who would have predicted are a patient's disease status typhoid or dengue. The algorithm used is C5.0 then the model is represented in the form of rules (rule). from the model established obtained accuracy greater than 84%.

Keywords : Dengue Haemorrhagic Fever, typhoid, C5.0 algorithm

1 PENDAHULUAN

1.1 Latar Belakang

Seiring dengan perkembangan zaman dan teknologi, ukuran *Database* saat ini sudah berkembang menjadi sangat besar ke dalam ukuran *terabyte*. Di dalam suatu tumpukan data pada *database* tersebut mungkin terdapat informasi-informasi tersembunyi yang penting atau akan menjadi penting pada saat-saat tertentu. Akan tetapi untuk menemukan hal tersebut sama saja seperti mencari jarum dalam tumpukan jerami. Oleh karena itu dapat disimpulkan bahwa untuk mengubah data mentah untuk dapat menjadi suatu informasi yang berguna bukan merupakan hal yang mudah.

Data mentah (*raw data*) sering dianggap tidak berguna dikarenakan ukurannya yang begitu besar sehingga sulit dianalisa. Oleh karena itu perlu dilakukan pengenalan pola dari data mentah tersebut. Solusinya adalah dengan *data mining*. Banyak organisasi di dunia telah menggunakan data mining untuk mencari dan menarik kesimpulan dari data yang mereka miliki.

Penggunaan *data mining* dilatar belakangi oleh banyaknya data yang diperoleh oleh berbagai institusi, perusahaan atau organisasi. Berlimpahnya data ini merupakan akumulasi data transaksi yang terekam selama 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 manajemen, pemasaran, 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. Salah satu masalah yang sering terjadi yaitu dalam mendiagnosis penyakit typhus dan demam berdarah, karena keduanya memiliki ciri-ciri dan gejala yang hampir sama sehingga sulit membedakannya, namun sebenarnya berbeda. Pada penelitian ini digunakan data *medical record* pasien suatu rumah sakit untuk penyakit demam berdarah dan typhus yang dibuat suatu modelan agar dapat diketahui faktor apa saja yang berpengaruh dalam mendiagnosis penyakit tersebut.

Problem seperti ini dalam *data mining* termasuk dalam jenis klasifikasi. Pada tugas akhir ini dibuat suatu model *decision tree* untuk memprediksi seorang pasien masuk dalam kategori typhus, DBD atau komplikasi keduanya dilihat dari data *medical record* pasien pada suatu rumah sakit berdasarkan ciri-ciri dan gejala pasien. Untuk membantu proses klasifikasi *data*, penulis menggunakan Algoritma yang sering digunakan untuk proses klasifikasi data dengan nilai akurasi yang tinggi yaitu Algoritma C5.0. Pemilihan algoritma ini dikarenakan Algoritma C5.0 tidak menggunakan vektor jarak untuk mengklasifikasikan obyek sehingga cocok untuk *data* observasi dengan atribut yang bersifat *numerik* ataupun atribut yang bernilai nominal yaitu bersifat *kategorik* dimana tiap nilainya tidak bisa dijumlahkan atau dikurangkan contohnya seperti bentuk, warna dan rasa. Selain itu algoritma ini juga menghasilkan *tree* dengan jumlah cabang per *node* bervariasi sehingga dapat membentuk pohon keputusan yang lebih efisien.

1.2 Rumusan Masalah

Berdasarkan uraian dari latar belakang yang telah dikemukakan di atas, permasalahan yang akan dibahas yaitu bagaimana mengklasifikasikan dan memprediksi penyakit Demam Berdarah dan Tipes yang diderita oleh pasien agar mendapatkan hasil dengan tingkat akurasi yang tinggi?

1.3 Tujuan

Penelitian ini bertujuan untuk membentuk *rule* guna mengklasifikasikan prediksi penyakit typhus atau DBD yang diderita oleh pasien berdasarkan ciri-ciri dan gejalanya dengan menggunakan teknik klasifikasi pohon keputusan dan algoritma C5.0 untuk menghasilkan rule yang seakurat mungkin.

1.4 Batasan Masalah

Untuk lebih memfokuskan penelitian dan menyederhanakan permasalahan, peneliti menentukan batasan masalah penelitian. Batasan yang digunakan dalam penelitian ini adalah:

1. Model dikembangkan untuk kasus klasifikasi tiga kelas yaitu typhus, DBD atau komplikasi keduanya.
2. Data yang digunakan untuk pengujian model adalah *dataset* kasus nyata yang didapatkan dari *medical record* pasien rumah sakit Sartika Asih.
3. Penelitian dilakukan dengan menggunakan metode pohon keputusan dan algoritma C5.0.

1.5 Metodologi Penyelesaian Masalah

Metodologi yang digunakan dalam menyelesaikan Tugas Akhir ini adalah sebagai berikut :

- a) Melakukan studi literatur mengenai algoritma C5.0
- b) Mengidentifikasi masalah ciri penyakit Demam Berdarah dan Typhus.
- c) Pengambilan data dari rumah sakit.
- d) Melakukan perancangan dan permodelan pada sistem yang akan diuji.
- e) Melakukan pengujian perangkat lunak dengan melihat performansi dan akurasi algoritma C5.0 terhadap klasifikasi kedua penyakit tersebut.
- f) *Testing* dan Analisis Hasil.
 - a. *Testing*
Melakukan pengujian sistem yang telah dibangun menggunakan data uji dengan berbagai skenario pengujian.
 - b. Analisis Hasil
Untuk membuktikan bahwa tingkat performansi sistem yang dibangun telah mencapai tingkat yang baik, maka dilakukan analisis terhadap performansi sistem sesuai dengan skenario pengujian yang dilakukan.
- g) Penyusunan laporan
Menyusun laporan dan dokumentasi terhadap penelitian yang telah dilakukan, serta membuat kesimpulan dari hasil analisis tersebut dengan mengikuti

ketentuan yang telah ditetapkan oleh institusi. Laporan yang dibuat antara lain adalah buku Tugas Akhir dan jurnal.

1.6 Sistematika Penulisan

Dalam penulisan Tugas Akhir ini dibagi menjadi beberapa bab yang meliputi hal-hal sebagai berikut :

1. Bab I Pendahuluan
Berisi latar belakang, tujuan, rumusan masalah, batasan masalah, tahapan penyelesaian masalah yang digunakan, serta sistematika penulisan yang memuat susunan penulisan Tugas Akhir ini.
2. Bab II Dasar Teori
Bab ini membahas tentang teori-teori yang mendukung yaitu, informasi mengenai penyakit demam berdarah dan typhus serta algoritma C5.0
3. Bab III Perancangan Sistem
Bab ini membahas tentang perancangan sistem untuk membuat sistem prediksi penyakit dengan algoritma C5.0.
4. Bab IV Implementasi dan Analisa Sistem
Bab ini membahas tentang pengujian sistem & analisis terhadap implementasi Algoritma C5 dalam memprediksi penyakit demam berdarah dan typhus
5. Bab V Kesimpulan dan Saran
Bab terakhir ini memberikan kesimpulan hasil penelitian pada Tugas Akhir yang telah dilakukan dan saran terhadap pengembangan ke depan.

5 KESIMPULAN DAN SARAN

5.1 Kesimpulan

Kesimpulan dari Tugas Akhir ini adalah :

1. Dengan menggunakan algoritma C5.0 nilai akurasi dari *data learn* dan *data test* yang di dapat lebih dari 80%. Hal ini menunjukkan bahwa algoritma C5.0 cukup layak untuk mengklasifikasikan penyakit demam berdarah dan typhus.
2. Tingkat keragaman *record* yang ada pada *data learn* yang digunakan dalam pembentukan pohon keputusan sangat mempengaruhi tingkat akurasi dari algoritma C5.0. oleh karena itu perlu dilakukan pemilihan *record* untuk digunakan sebagai *data learn* untuk mendapatkan akurasi yang lebih baik.
3. Waktu komputasi dari sistem berbanding lurus dengan jumlah dataset yang digunakan terutama jika data set diatas 10000.
4. *Attribut* widal merupakan *attribut* yang paling berpengaruh dalam proses klasifikasi penyakit demam berdarah dan typhus berdasarkan *dataset*.

5.2 Saran

Saran-saran untuk pengembangan lebih lanjut diantaranya :

1. Perlu adanya metode untuk pemilihan *record* yang tepat untuk menjadi *data learn* untk menghasilkan akurasi yang maksimal.
2. Perlu adanya metode untuk pemilihan *attribut* yang lebih baik dalam pembentukan *rule*.

DAFTAR PUSTAKA

- [1] Anugrah, Fitran. Imbalance Class: <http://elreg-02.blogspot.com/2010/01/teknik-boosting.html>
- [2] Ariadni, Ratih. Arieshanti, Isye. Implementasi Pohon Keputusan Untuk Klasifikasi Data Dengan Nilai Fitur Yang Tidak Pasti. Teknik Informatika, Fakultas Teknologi Informasi, Institut Teknologi Sepuluh Nopember (ITS). Surabaya.
- [3] Basuki, Achmad. Syarif, Iwan. 2003. Decision Tree. Politeknik Elektronika Negeri Surabaya
- [4] Bertalya. 2009. Klasifikasi Pohon Keputusan. Universitas Gunadarma.
- [5] Billing, Jeffrey J. 2002. Learning Classifiers from Medical Data. The Massachusetts Institute of Technology.
- [6] Demam berdarah atau Typhus: <http://pkukertosono.blogspot.com/2011/07/demam-berdarah-atau-typhus.html>
- [7] Ernawati lin, 2008. Prediksi Keaktifan Studi mahasiswa dengan Algoritma C5.0 dan K-Nearest Neighbor[Tesis]. Program Pasca Sarjana Institut Pertanian Bogor
- [8] Prawitasari, Anita. Classification and by Prediction DBD Tipes Algorithms Decision Tree. Universitas Gunadarma.
- [9] Hidayanto, Achmad N dkk. Perbandingan Algoritma Klasifikasi Dalam Pendeteksian Penyakit Kanker. Makalah Fakultas Teknik Komputer, Universitas Indonesia.
- [10] Kusnawi. 2007. Pengantar Solusi Data Mining. STMIK AMIKOM. Yogyakarta.
- [11] Shaufiah. Data Bagian 2. KBK RPL dan Data Fakultas Informatika, IT Telkom.
- [12] Suryana, Akuh. Penyakit Typhes (Thyphoid): <http://akuhsuryana.wordpress.com/penyakit-types-typhoid/>
- [13] Suyanto.: *Artificial Intelligence: Searching, Reasoning, Planning and Learning*. Penerbit Informatika, Bandung, ISBN : 9786028758208 (2011).
- [14] Ramadoni. Mengenai DHF, sudah tahukah anda???: <http://www.ramadoni.com/2009/08/mengenai-dhf-sudah-tahukah-anda/>
- [15] Yusuf W, Yogi. 2007. Perbandingan Performansi Algoritma Decision Tree C5.0, CART, dan Chaid: Kasus Prediksi Status Resiko Kredit Di Bank X. Paper Jurusan Teknik Industri, Universitas katolik Parahyangan