

PREDIKSI PENYAKIT JANTUNG DENGAN ALGORITMA CLASSIFICATION AND REGRESSION TREE

Nita Apriliani Puteri¹, Warih Maharani², Mahmud Dwi Suliyo³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Penyakit Jantung merupakan salah satu penyakit yang dapat sangat mematikan di dunia. Untuk menekan persentase kematian akibat penyakit jantung maka dibutuhkan sebuah sistem yang dapat memprediksi penyakit jantung.

Algoritma CART adalah metode klasifikasi untuk menentukan prediksi yang dapat menangani himpunan data yang mempunyai jumlah besar dan variabel yang sangat banyak. Algoritma CART akan menghasilkan suatu pohon klasifikasi jika variabel responnya kategorik, dan menghasilkan pohon regresi jika variabel responnya kontinu. Dan akan menyeleksi beberapa variabel dan interaksi antar variabel yang paling penting dalam menentukan hasil variabel dependennya. Ciri khas algoritma CART adalah akan menghasilkan kelompok data yang akurat dengan hasil yang mudah diinterpretasikan dan lebih cepat dalam hal perhitungannya.

Dalam Tugas Akhir ini akan dikupas mengenai langkah kerja algoritma CART pada prediksi penyakit jantung untuk mengetahui hasil prediksi penyakit jantung, dan mengetahui variabel yang berpengaruh dalam membentuk prediksi penyakit jantung.

Hasil pengujian menunjukkan bahwa total seluruh variabel yang berpengaruh dalam menghasilkan prediksi penyakit jantung berjumlah delapan buah, yaitu umur, total kolesterol, HDL kolesterol, systolic, diastolic, status diabetes, status merokok, dan jenis kelamin dan parameter penentu utamanya adalah parameter umur. Adanya pengaruh antara kombinasi jumlah data training yang bernilai positif dan negatif, jumlah data training menentukan perbedaan tingkat diferensiasi parameter yang terjadi. Prediksi penyakit jantung dengan algoritma CART ini dapat menghasilkan akurasi tertinggi mencapai 77%.

Kata Kunci : prediksi, jantung, algoritma, CART, pohon, klasifikasi.

Telkom
University

Abstract

Heart disease is one of disease that can be deadly in the world. To decrease the percentage of deaths from heart disease it needed a system that can predict heart disease.

CART algorithm is a classification method for determining the set of predictions that can handle large amounts of data and variables. CART algorithm will produce a classification tree if the response was categorical variables, and produce regression tree if the response variable is continuous. CART algorithm will select some variables and its interactions are most important to determine the outcome of the dependent variable. Characteristic CART algorithm is the data will produce accurate results faster and more easily interpreted in terms of the calculation.

In this final project will be discussed about prediction of heart disease by using CART Algorithm to determine the outcome prediction of heart disease, and know the variables that can affect the prediction of heart disease.

The experimental results show that the total of all the variables that can affect the result in the prediction of heart disease were eight variables, there are age, total cholesterol, HDL cholesterol, systolic, diastolic, diabetes status, smoking status, and gender, and the main parameter is age. The other factor that show influence of the result are training's data combination which is positive and negative, the total amount of training data to determine the differentiation of parameters. Prediction of heart disease by CART algorithm can produce the highest accuracy that can be reached is 77%

Keywords : prediction, heart, algorithm, CART, tree, classification.

1. Pendahuluan

1.1 Latar belakang masalah

Tingginya penyebab kematian manusia salah satunya disebabkan oleh penyakit jantung. Penyakit jantung merupakan salah satu penyakit yang menjadi pusat ketakutan masyarakat dunia. Berdasarkan Jurnal Kedokteran, 2003 diungkapkan bahwa penyakit jantung merupakan penyebab disabilitas dan kerugian ekonomis yang tertinggi dibanding penyakit lain[12]. Salah satu contoh nyata yang ada terkait penyakit jantung ini, di negara berkembang dari tahun 1990 sampai 2020, angka kematian akibat penyakit jantung koroner akan meningkat 137 % pada laki-laki dan 120% pada wanita sedangkan di negara maju peningkatannya lebih rendah yaitu 48% pada laki-laki dan 29% pada wanita[12]. Di Indonesia, dilaporkan bahwa penyakit jantung merupakan penyebab kematian nomor satu. Salah satu langkah yang dapat ditempuh untuk meminimalisir atau menormalkan angka kematian manusia disebabkan penyakit jantung antara lain dengan membuat sebuah prediksi penyakit jantung, agar masyarakat dapat waspada, dan mulai memperbaiki pola hidupnya agar mencegah terenggutnya nyawa akibat penyakit jantung.

Beberapa metode dapat ditempuh untuk menciptakan sebuah prediksi. Namun, prediksi yang dibutuhkan dalam kasus penyakit jantung ini adalah sebuah prediksi yang akurat dengan hasil yang mudah diinterpretasikan dan lebih cepat dalam hal perhitungan. Selain itu, terdapat beberapa variabel yang dapat mempengaruhi prediksi penyakit jantung ini sehingga dibutuhkan sebuah metode prediksi yang dapat menangani himpunan data yang mempunyai jumlah besar dan variabel yang sangat banyak, metode tersebut adalah algoritma CART.

Algoritma CART (*Classification and Regression Trees*) memiliki beberapa kelebihan antara lain lebih mudah diinterpretasikan, lebih akurat dan lebih cepat penghitungannya dan diterapkan untuk himpunan data yang mempunyai jumlah besar, variabel yang sangat banyak dan dengan skala variabel campuran melalui prosedur pemilahan biner. Menurut Yohannes dan Webb (1999), tingkat kepercayaan yang bisa digunakan dalam mengklasifikasikan data baru pada CART adalah akurasi yang dihasilkan oleh pohon klasifikasi yang murni dibentuk dari data yang mempunyai kesamaan kondisi (*data learning*) [2]. Algoritma CART biasa digunakan untuk melakukan klasifikasi namun dalam kasus ini algoritma CART akan digunakan untuk memprediksi penderita penyakit jantung.

1.2 Perumusan Masalah

Beberapa masalah yang ingin diselesaikan dari hasil penelitian Tugas Akhir ini adalah

1. Bagaimana hasil prediksi penyakit jantung yang dihasilkan oleh algoritma CART?
2. Apa variabel utama yang berpengaruh dalam membentuk prediksi penyakit jantung?

1.3 Batasan Masalah

Dalam penelitian Tugas Akhir ini, dibutuhkan sejumlah dataset, baik untuk data *training* maupun data *test*. Data yang akan digunakan sebagai input sistem prediksi penyakit jantung berasal dari rekam status Rumah Sakit. Data yang akan diambil adalah data yang memiliki karakteristik yang berbeda-beda, seperti umur yang memiliki karakteristik data berupa data kategorik, jenis kelamin merupakan data dengan karakteristik kategorik, tekanan gula darah merupakan data dengan karakteristik data kategorik, dan lain-lain.

1.4 Tujuan

Penelitian Tugas Akhir ini dilakukan dengan tujuan sebagai berikut

1. Mengetahui hasil prediksi penyakit jantung dengan algoritma *Classification And Regression Tree*.
2. Mengetahui variabel utama yang berpengaruh dalam prediksi penyakit jantung.

1.5 Metodologi Penyelesaian Masalah

Beberapa tahap penyelesaian masalah yang akan dilakukan meliputi

1. Studi Literatur
Pencarian informasi terkait metode algoritma yang digunakan untuk prediksi penyakit Jantung ini, baik dari buku, paper, maupun literatur lain dari internet. Pencarian informasi terkait penyakit jantung diperoleh dari seorang dokter spesialis penyakit jantung, yakni dr. Guntur dari Rumah sakit Al-Islam, Bandung.
2. Tahap Perancangan
Pada tahap perancangan ini, menetapkan variabel yang akan digunakan untuk prediksi penyakit jantung ini, menentukan data dan informasi yang dibutuhkan untuk penelitian ini, yakni dataset *training* dan data *test*, serta menentukan rancangan sistem prediksi penyakit jantung yang akan dibuat yakni berupa *flowchart* sistem.
3. Pengumpulan data
Pencarian serta pengumpulan data dan informasi yang dibutuhkan untuk penelitian prediksi penyakit jantung ini, baik dari nara sumber, dr. Guntur ataupun data status pasien di Rumah Sakit Al-Islam, Bandung.
4. Implementasi
Pembuatan sistem prediksi penderita penyakit jantung. Sistem prediksi penyakit jantung ini akan diimplementasikan dengan menggunakan Matlab.

Sedangkan untuk pengolahan data *training* dengan menggunakan tools Spawner dan Microsoft Excel 2007.

5. Pengujian dan Analisis Hasil
Melakukan beberapa *test* dan pengujian terhadap sistem yang telah dibuat, dan akan dibandingkan keakuratannya dengan data *test* berupa data status pasien dari Rumah Sakit Al Islam, Bandung.
6. Pembuatan Laporan
Mendokumentasikan seluruh penelitian yang telah dilakukan serta menarik kesimpulan berdasarkan penelitian ini.

1.6 Sistematika Penulisan Tugas Akhir

Sistematika Tugas Akhir ini terdiri dari lima bab utama, yaitu

BAB I. PENDAHULUAN

Berisi latar belakang, perumusan masalah, batasan masalah, tujuan, metodologi penyelesaian masalah dan sistematika (*outline*) Tugas Akhir.

BAB II. LANDASAN TEORI

Berisi penjelasan singkat mengenai konsep-konsep yang mendukung dikembangkannya sistem ini. Konsep-konsep yang digunakan untuk mendukung sistem ini adalah konsep Framingham Heart Study sebagai dasar penentu prediksi penyakit jantung, pendekatan dan jenis metode klasifikasi, dasar algoritma *Classification And Regression Tree* dan langkah kerja algoritma *Classification And Regression Tree*.

BAB III. ANALISIS KEBUTUHAN DAN PERANCANGAN SISTEM

Berisi mengenai analisis kebutuhan perangkat lunak dan perangkat keras yang akan dibangun serta rincian perancangan sistem untuk menentukan prediksi penyakit jantung dengan algoritma *Classification And Regression Tree*.

BAB IV. IMPLEMENTASI DAN ANALISIS HASIL PENGUJIAN

Berisi rincian pengujian terhadap sistem yang telah dibangun disertai analisis terhadap hasil pengujian sistem tersebut. Pada bab ini akan dilakukan analisis terhadap hasil prediksi penyakit jantung dengan algoritma *Classification And Regression Tree* serta variabel yang berpengaruh dalam menentukan prediksi penyakit jantung dengan algoritma *Classification And Regression Tree*.

BAB V. KESIMPULAN DAN SARAN

Berisi tentang kesimpulan yang didapat dari pelaksanaan Tugas Akhir ini dan saran-saran yang diperlukan untuk perbaikan maupun pengembangan lebih lanjut.

5. Penutup

5.1 Kesimpulan

Berdasarkan hasil implementasi dan pengujian yang telah dilakukan pada bab 4, maka dapat ditarik kesimpulan bahwa

1. Akurasi terbesar pada data *training* sejumlah 800 data dengan jumlah data *training* yang bernilai positif 211 mencapai 76% sedangkan pada data *training* sejumlah 1000 data dengan jumlah data *training* yang bernilai positif 330 data yakni sebesar 77%.
2. *Pruning* level berpengaruh pada hasil akurasi, *tree* yang dihasilkan setelah *pruning* akan lebih optimal dan parameter utama penentu prediksi penyakit jantung dengan menggunakan algoritma CART adalah parameter umur.
3. Kombinasi jumlah data *training* yang bernilai positif dengan negatif berpengaruh pada akurasi yang dihasilkan.

5.2 Saran

Untuk pengembangan sistem yang lebih baik, terdapat beberapa hal yang dapat dikembangkan dari sistem ini, yakni

1. Memperluas jangkauan data dengan menghilangkan batasan data yang terdapat pada sistem ini
2. Melakukan lebih banyak pengujian dengan jumlah diferensiasi data yang lebih besar
3. Meningkatkan akurasi prediksi penyakit jantung ini dengan menggabungkan metode lain

Daftar Pustaka

- [1] Abdul, Mohamad Kadir. 2010. *Perbandingan Performansi Algoritma Decision Tree CART dan Chaid*. Institut Teknologi Bandung : Bandung.
- [2] Aprian, Krisan Widagdo. 2010. *Pembentukan Pohon Klasifikasi Biner Dengan Algoritma CART (Classification And Regrsson Trees) Studi Kasus Penyakit Diabetes Suku Pima Indian*. Semarang.
- [3] Janarko, Arief, Sulastri dan Rara Sri Artati R. 2010. *Perbandingan Performansi Algoritma Nearest Neighboar dan SLID untuk prediksi Kinerja Akademik Mahasiswa Baru (Studi Kasus : Data Akademik Mahasiswa Fakultas Teknologi Informasi UNISBANK)*. Surabaya.
- [4] Pang, dkk. 2006. *Introduction Data Mining*. Pearson Education : Boston.
- [5] Parthiban, Latha dan R. Subramanian. 2007. *Intelligence Heart Disease Prediction System using CANFIS and Genetic Algoritm*. International Journal of Biological and Life Sciences 3-3.
- [6] Supriyono, Mamat. 2008. *Thesis : Faktor-faktor risiko yang berpengaruh terhadap kejadian penyakit jantung koroner pada kelompok usia < 45 tahun (Study Kasus Di RSUP Dr. Kariadi Dan RS Telogorejo Semarang)*. Semarang
- [7] Shi, Haijian. 2006. *Thesis : Best-first Decision Tree Learning*. University of Waikato. NewZealand
- [8] Timofeev, Roman. 2004. *Thesis : Classification And Regression Trees(CART) Theory And Applications*. 2004. CASE - Center of Applied Statistics and Economics Humboldt University. Berlin
- [9] Widjanarko, Bambang Otok dan Sumarmi. 2009. *Bagging CART Pada Klasifikasi Anak Putus Sekolah*. Surabaya.
- [10] Wilson PWF, et al. *Prediction od coronary heart desease using risk factor categories*. Circulation 1998;97:1837-47.
- [11] Yusuf, Yogi W. 2007. *Perbandingan Performansi Algoritma Decision Tree c5.0, CART, Dan Chaid : Kasus Prediksi Status Risiko Kredit Di Bank X*. Bandung.
- [12] Zulkifli, B. 2011. *Penyakit Jantung Koroner*. Sumatra Selatan.