

CASE-BASED REASONING UNTUK DIAGNOSIS PENYAKIT JANTUNG MENGUNAKAN METODE MINKOWSKI DISTANCE

CASE-BASED REASONING FOR HEART DISEASES DIAGNOSIS USING MINKOWSKI DISTANCE METHOD

Eka Wahyudi¹, Novi Indah Pradasari²

Politeknik Negeri Ketapang¹²

ekawahyudi_algebra@ymail.com¹, noviindah.ti@gmail.com²

ABSTRACT

Case Based Reasoning (CBR) is a computer system that used for reasoning old knowledge to solve new problems. It works by looking at the closest old case to the new case. This research attempts to establish a system of CBR for diagnosing heart disease. The diagnosis process is done by inserting new cases containing symptoms into the system, then the similarity value calculation between cases uses the minkowski distance similarity. Case taken is the case with the highest similarity value. If a case does not succeed in the diagnosis or threshold <0.80 , the case will be revised by experts. Revised successful cases are stored to add the system knowledge. Method with the best diagnostic result accuracy will be used in building the CBR system for heart disease diagnosis. The test results using medical records data validated by expert indicate that the system is able to recognize diseases heart using minkowski distance similarity correctly of 100%. Using minkowski get accuracy of 100%.

Keywords : Case-Based Reasoning, Minkowski Distance Similarity.

ABSTRAK

Case Based Reasoning (CBR) merupakan sistem penalaran komputer yang menggunakan pengetahuan lama untuk mengatasi masalah baru. CBR memberikan solusi terhadap kasus baru dengan melihat kasus lama yang paling mendekati kasus baru. Sistem yang dibangun dalam penelitian ini adalah sistem CBR untuk diagnosis penyakit Jantung. Proses diagnosis dilakukan dengan memasukkan kasus baru yang berisi fitur-fitur yang akan didiagnosis ke dalam sistem, kemudian melakukan proses perhitungan nilai similaritas antara kasus baru dengan basis kasus minkowski distance. Kasus yang diambil adalah kasus dengan nilai similaritas tertinggi. Jika suatu kasus tidak berhasil didiagnosis atau threshold $<0,8$, maka akan dilakukan revisi kasus oleh pakar. Kasus yang berhasil direvisi disimpan kedalam sistem untuk dijadikan pengetahuan baru. Metode dengan akurasi hasil diagnosis terbaik akan digunakan untuk membangun sistem CBR diagnosis penyakit jantung. Hasil pengujian rata-rata sistem untuk melakukan diagnosis awal terhadap penyakit jantung menggunakan data rekam medik pasien penyakit jantung, menunjukkan bahwa sistem mampu mengenali penyakit jantung menggunakan metode minkowski distance similarity secara benar sebesar 100%. Serta perhitungan tingkat akurasi sebesar 100%.

Kata kunci : Case-Based Reasoning, Minkowski Distance Similarity.

PENDAHULUAN

Penyakit jantung merupakan salah satu penyakit yang berbahaya dan penyebab kematian nomor satu di dunia jika tidak ditangani dengan baik. Di Indonesia khususnya pada provinsi Yogyakarta, menurut data rekam medis RSUP Dr. Sardjito Yogyakarta jumlah penderita penyakit jantung dari tahun ketahun terus meningkat. Dua puluh lima (25) besar

jumlah pasien penderita penyakit jantung pada tahun 2012 sebanyak 807 orang, tahun 2013 bertambah menjadi 891 orang dan pada tahun 2014 mulai dari bulan Januari sampai pada bulan Nopember berjumlah 918 orang (Instalasi Catatan Medis, 2014).

Untuk mendiagnosis pasien yang menderita penyakit jantung dapat diketahui dari gejala-gejala yang muncul atau dialami oleh pasien dan faktor resiko yang dirasakan

pasien. Penanganan pada pasien penyakit jantung dilakukan oleh tenaga ahli yaitu dokter spesialis penyakit jantung, namun karena keterbatasan tenaga ahli mengakibatkan lambatnya penanganan pasien, sehingga dibutuhkan suatu alat yang dapat membantu untuk diagnosis pasien penyakit jantung.

Kasus-kasus yang tersimpan dalam rekam medis mengenai diagnosis penyakit jantung oleh tenaga ahli untuk menentukan jenis penyakit jantung yang diderita oleh pasien dapat digunakan kembali sebagai acuan untuk menentukan jenis penyakit jantung ketika ada kasus baru. Pemanfaatan kasus yang telah terjadi sebelumnya atau kasus lama dikenal secara umum dengan istilah penalaran berbasis kasus atau *case based reasoning* (CBR).

Dalam CBR sendiri ada beberapa proses yaitu *retrieve*, *reuse*, *revise* dan *retain*. Pada fase *retrieve* ada banyak metode yang di gunakan untuk pengambilan kasus lama yang relevan dengan kasus baru. Fase *retrieve* merupakan salah satu tahap penting dalam siklus CBR. Beberapa metode *retrieve* adalah dengan menggunakan metode *nearest neighbor similarity*, *minkowski distance similarity* dan *euclidean distance similarity* (Aamodt dan Plaza, 1994).

Fase *revise* merupakan bagian dari adaptasi sistem terhadap kasus yang belum berhasil didiagnosis. Dimana kasus baru tidak diperoleh kasus yang identik atau nilai similaritas tertinggi $< 0,8$, maka solusi yang disarankan adalah menyerahkan kasus tersebut kepada pakar.

Penelitian ini menyajikan pengembangan sebuah sistem CBR untuk diagnosis penyakit jantung dengan menggunakan metode pengukuran similaritas *minkowski distance similarity* yang mengakomodir tingkat keyakinan pakar.

METODE

Metode yang digunakan dalam penelitian ini adalah pengembangan sebuah sistem CBR untuk diagnosis penyakit jantung dengan menggunakan metode pengukuran similaritas *minkowski distance*

Tiap atribut memiliki bobot yang

similarity yang mengakomodir tingkat keyakinan pakar. Secara garis besar, pengukuran similaritas yang digunakan dalam penelitian ini mencakup *Local Similarity*, Pengukuran Tingkat Keyakinan, dan *Global Similarity*.

1. Case-Based Reasoning

Case-based reasoning (CBR) merupakan metode penyelesaian masalah yang menggunakan pengetahuan akan pengalaman terdahulu untuk memecahkan permasalahan yang baru (Pal, S.K., dan Shiu, S.C.K., 2004). Sebuah permasalahan diselesaikan dengan mencari sebuah kasus lama yang identik, jika ditemukan maka solusi dari keduanya juga identik. Namun apabila tidak ditemukan, sistem akan mencari kasus lama yang memiliki kemiripan tertinggi dan namun perlu dilakukan adaptasi sehingga diperoleh solusi yang sesuai dengan permasalahan baru tersebut. Alur CBR meliputi :

- Retrieve*, mencari kasus terdahulu yang serupa dengan permasalahan
- Reuse*, menyalin atau menggabungkan solusi dari kasus terdahulu
- Revise*, mengadaptasi solusi dari kasus terdahulu menjadi solusi permasalahan yang baru
- Retain*, solusi yang baru tersebut divalidasi

2. Proses Retrieval

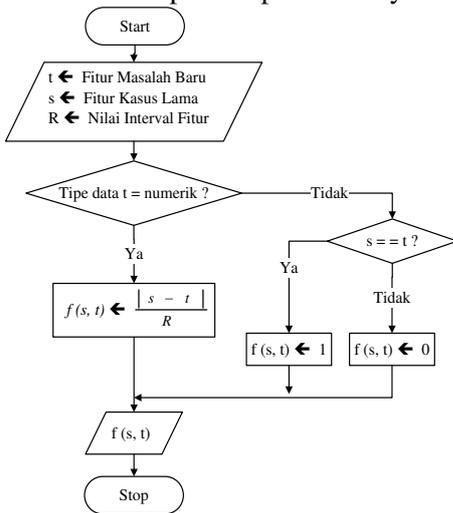
CBR didasarkan pada hipotesa bahwa solusi permasalahan-permasalahan sebelumnya dapat membantu penyelesaian permasalahan terkini, sepanjang terdapat kemiripan diantara mereka. Pengukuran kemiripan dilakukan dengan membandingkan fitur yang ada pada kasus dengan fitur sejenis yang ada pada basis kasus. Suatu kasus disebut identik dengan kasus lain apabila nilai similaritasnya sama dengan satu, dan dikatakan mirip jika nilainya dibawah satu. Secara garis besar pengukuran similaritas terdiri dari pengukuran pengukuran similaritas lokal antar atribut kasus dan similaritas global antar kasus. Kemiripan antar atribut ini disebut sebagai similaritas lokal (Jha dkk, 2013).

berbeda untuk setiap jenis solusi/penyakit

yang nilainya ditentukan oleh pakar.

3. Similaritas Lokal

Similaritas lokal menunjukkan keserupaan antara atribut permasalahan terhadap atribut yang sama dari sebuah kasus. Perhitungan similaritas lokal dihitung berdasarkan tipe data pada fiturnya.



Gambar. 1 Perhitungan Similaritas Lokal

(Sumber: Dimodifikasi dari Jha, M.K., Pakhira, D., dan Chakraborty, B., 2013)

4. Tingkat Keyakinan

Tingkat keyakinan adalah pencerminan tingkat kepastian yang diinginkan oleh pengukur setelah memutuskan tidak akan melakukan pengukuran yang sangat banyak (Nurdiansyah, Y., dan Hartati, S., 2014). Pengukuran tingkat keyakinan yang digunakan dalam penelitian ini adalah tingkat keyakinan pakar dan ting tingkat keyakinan kasus baru terhadap basis kasus. Tingkat keyakinan pakar ditentukan oleh pakar berdasarkan gejala dan faktor resiko yang dialami pasien. Untuk menghitung tingkat keyakinan kasus baru terhadap basis kasus menggunakan rumus.

$$\mu_{(T,S)} = \frac{J(S,T)}{J(T)}$$

Keterangan:

$\mu_{(T,S)}$: Tingkat keyakinan antara kasus T (*target case*) dan S (*source case*)

$J(S,T)$: Banyaknya fitur yang terdapat dalam *target case* yang muncul pada fitur *source case*

$J(T)$: Banyaknya fitur yang terdapat dalam *target case*

5. Similaritas Global

Similaritas global digunakan untuk menghitung keserupaan antar permasalahan dengan kasus yang tersimpan dalam basis kasus. Metode pengukuran similaritas global yang digunakan dalam penelitian ini adalah *minkowski distance similarity* (Hastie, T., Tibshirani, R., dan Friedman, J., 2009) dengan pembobotan dan telah di modifikasi dengan menambahkan faktor tingkat keyakinan menggunakan rumus.

$$SimMD(S,T) = \left(\frac{\sum_{i=1}^n (w_{i,p(S)})^3 * |f_i(S_i, T_i)|^3}{\sum_{i=1}^n (w_{i,p(S)})^3} \right)^{1/3} * P(S) * \frac{J(S_i, T_i)}{J(T_i)}$$

Keterangan:

$SimMD(S,T)$: Similaritas global antara kasus T (*target case*) dan S (*source case*)

N : Banyaknya fitur yang ada

$w_{i,p(S)}$: Nilai bobot fitur ke-i pada penyakit dari *source case*

R : Faktor minkowski (integer positif) (dalam hal ini r=3)

$f_i(S_i, T_i)$: Kesamaan fitur ke-i dari *source case* dan *target case* / fungsi similaritas lokal

S_i : Fitur ke-i yang ada dalam *source case*

T_i : Fitur ke-i yang ada dalam *target case*

$P(S)$: Persentase tingkat keyakinan pakar terhadap suatu kasus dalam *source case*

$J(S_i, T_i)$: Banyaknya fitur yang terdapat dalam *target case* yang muncul pada fitur *source case*

$J(T_i)$: Banyaknya fitur yang terdapat dalam *target case*

6. Fase Adaptasi

Proses retrieval menghasilkan sebuah kasus lama dengan similaritas tertinggi. Apabila dari hasil proses *retrieval* tidak diperoleh kasus yang identik maka solusi yang disarankan oleh sistem adalah solusi dari kasus terdahulu yang memiliki nilai similaritas tertinggi. Solusi tersebut terkadang tidak selalu tepat sehingga perlu dilakukan proses adaptasi.

Beberapa teknik yang dapat digunakan untuk adaptasi (Mergio, J.M., dan Casanovas, M., 2008) yaitu :

- Tidak ada adaptasi (*Null Adaptation*)
- Penyesuaian atau pengaturan parameter
- Reinstantiation*
- Derivational replay*

e. Perbaiki model terpandu

Model adaptasi yang digunakan dalam penelitian ini yaitu dengan menetapkan suatu nilai ambang (*threshol*). Pengecekan nilai similaritas tertinggi dari hasil proses *retrieval* dilakukan. Apabila nilainya $\geq 80\%$ (nilai *threshol*) maka solusi dari kasus dengan similaritas tertinggi ditetapkan sebagai solusi dari permasalahan dan jika nilainya $< 80\%$ maka permasalahan akan disimpan untuk kemudian dilakukan revisi oleh pakar.

7. Data dan Metode Pengujian

Untuk menguji penggunaan CBR untuk diagnosis penyakit jantung menggunakan metode *minkowski distance similarity* digunakan data rekam medik pasien rawat inap penderita penyakit jantung yang diperoleh dari Instalasi Catatan Medik RSUP Dr Sardjito Yogyakarta.

Tabel 1. Rekapitulasi Data Rekam Medik Untuk Basis Kasus

Jenis Penyakit	Jumlah Data	Data Basis Kasus (70%)	Data Uji (30%)
Gagal Jantung Akut	40	28	12
Jantung Koroner	40	28	12
Jantung Hipertensi	40	28	12
Gagal Jantung Kronik/Congestive	40	28	12
Jantung Katup/Endocarditis	16	11	4,8
Jantung Perikarditif	16	11	4,8
	192	134	58

Sumber : Instalasi Catatan Medik, 2014

HASIL DAN PEMBAHASAN

Perancangan CBR untuk diagnosis penyakit jantung menggunakan metode *minkowski distance similarity* ini diawali dengan analisis kebutuhan system, selanjutnya dilanjutkan dengan merancang desain interface aplikasi, lalu diimplementasikan dalam pembuatan system (coding) metode yang digunakan, dan selanjutnya dilakukan pengujian kepada data rekam medik yang akan menggunakan sistem CBR untuk diagnosis penyakit jantung menggunakan metode *minkowski distance similarity*. Berikut hasil dari penggunaan CBR untuk diagnosis penyakit

jantung menggunakan metode *minkowski distance similarity*.

1. Hasil Pengujian

Hasil pengujian dari keseluruhan data yang diujikan didapati dua permasalahan yang memiliki nilai similaritas dibawah 80% sehingga tidak dapat digolongkan ke dalam jenis penyakit jantung meskipun kasus dengan similaritas tertinggi memiliki diagnosis yang sesuai dengan diagnosis permasalahan. Rekapitulasi hasil pengujian dari keseluruhan data disajikan sebagai berikut:

Tabel 2. Rekapitulasi Hasil Pengujian

Jenis Penyakit	Jumlah Data Uji	Jumlah Diagnosa Benar
Gagal Jantung Akut	12	12
Jantung Koroner	12	12
Jantung Hipertensi	12	12
Gagal Jantung Kronik	12	12
Jantung Katup	5	5
Jantung Perikarditif	5	5
JUMLAH	58	58

2. Pembahasan

Evaluasi hasil pengujian sistem dalam mendiagnosis penyakit jantung dilakukan dengan menghitung sensitivitas dan akurasi (Witten, I.H., dan Frank, E., 2005). Evaluasi penting dilakukan untuk mengetahui apakah sistem yang dibuat layak diterapkan dalam mendiagnosis penyakit jantung. Untuk menghitung sensitivitas dan akurasi menggunakan rumus.

$$Sensitivitas = \frac{T_P}{T_P + F_N}$$

$$Akurasi = \frac{T_P + T_N}{T_P + F_P + T_N + F_N}$$

Keterangan:

- T_P : Banyaknya hasil diagnosis positif/benar untuk data uji positif
- F_P : Banyaknya hasil diagnosis negatif/salah untuk data uji positif
- T_N : Banyaknya hasil diagnosis positif/benar untuk data uji negatif
- F_N : Banyaknya hasil diagnosis negatif/salah untuk data uji negatif

Rekapitulasi *confusion matrix* hasil pengujian dan Rekapitulasi hasil perhitungan jenis pengukuran pengujian sebagai berikut:

Tabel 3. Confusion Matrix Hasil Pengujian

Jenis Penyakit	Jumlah Diagnosa Benar	
	Penyakit Jantung	Bukan Penyakit Jantung
Penyakit Jantung	58 (TP)	0 (FP)
Bukan Penyakit Jantung	0 (FN)	0 (TN)
TOTAL	68	0

Tabel 4. Rekapitulasi Hasil Perhitungan Jenis Pengukuran Pengujian

Jenis Pengukuran	Motode Minkowski Distance
Sensitivitas	100%
Akurasi	100%

PENUTUP

Berdasarkan penelitian dan hasil pengujian sistem maka dapat disimpulkan bahwa :

1. Penelitian ini menghasilkan sistem CBR untuk diagnosis penyakit jantung dengan memperhitungkan kedekatan antara permasalahan baru dan kasus lama berdasarkan fitur usia, fitur jenis kelamin, fitur gejala dan fitur faktor resiko dan dengan mengakomodasi bobot fitur kasus dan tingkat keyakinan.
2. Hasil pengujian terhadap data uji penyakit jantung menunjukkan bahwa sistem mampu mengenali penyakit jantung menggunakan metode *minskowski distance* secara benar masing-masing sebesar 100%.
3. Hasil pengujian data uji penyakit

jantung menunjukkan bahwa dengan nilai treshold similaritas global sebesar 80, sistem memiliki unjuk kerja dengan tingkat akurasi metode *minkowski distance similarity* sebesar 100%.

DAFTAR PUSTAKA

- Aamodt, A., dan Plaza, E (1994). Case Based Reasoning: Foundation Issues Methodological Variations. and Sistem Approaches, *AI Communication IOS Press*. 7, 1, 39-59.
- Instalasi Catatan Medis. (2014). *Data Rekam Medis*. RS. Dr. Sardjito Yogyakarta. Yogyakarta.
- Pal, S.K., dan Shiu, S.C.K.. (2004). *Fondation of Soft Case-Based Reasoning*. John Willey and Sons, Inc. New Jersey.
- Hastie, T., Tibshirani, R., dan Friedman, J. (2009). *The Element of Statistical Learning : Data Mining, Inference, and Prediction, Springer Series in Statistic*, 2. Springer-Verlag, Inc. New York.
- Jha, M.K., Pakhira, D., dan Chakraborty, B. (2013). Diabetes Detection and Care Applying CBR Techniques. *IJSCE*, 6, 2. 132-137.
- Mergio, J.M., dan Casanovas, M. (2008). The Induced Minkowski Ordered Weighted Averaging Distance Operator, ESTYLF08, Cuencas Mineras (Mieres-Langreo). *Congreso Espanol sobre Tecnologiasy Logica Fuzzy*. pp 35-41.
- Nurdiansyah, Y., dan Hartati, S. (2014). *Case-Based Reasoning untuk Pendukung Diagnosa Gangguan pada Anak Autis*. Thesis. Prodi S2/S3 Ilmu Komputer, Universitas Gadjah Mada, Yogyakarta.
- Witten, I.H., dan Frank, E. (2005). *Data Mining : Practical Machine Learning Tools and Techniques*, 2. Morgan Kaufmann Publisher. San Fransisco.