

Diagnosa Penyakit Kulit Menggunakan Case Based- Reasoning dan Self Organizing Maps

Fhatiah Adiba¹
Nurul Mukhlisah Abdal²
Andi Akram Nur Risal³

Universitas Negeri Makassar^{1,2,3}

Email: adibafhatiah@unm.ac.id¹

Abstract. This study aims to compare the results of the accuracy and speed of the system in diagnosing skin diseases using the case based reasoning (CBR) method with the indexing method and without using indexing. Self-organizing maps (SOM) are used as an indexing method and the process of finding similarity values uses the nearest neighbor method. Testing is done with two scenarios. The first scenario uses CBR without indexing self-organizing maps, the second scenario uses CBR with indexing self-organizing maps. The accuracy of the diagnosis of skin diseases at a threshold ≥ 80 for CBR without indexing self-organizing maps is 93.46% with an average retrieve time of 0.469 seconds while CBR testing using SOM indexing is 92.52% with an average retrieve time of 0.155 seconds. The results of comparison of CBR methods without using show higher results than using SOM indexing, but the process of retrieving CBR using SOM is faster than not using indexing.

Keywords: case based reasoning (CBR), self-organizing maps (SOM), nearest neighbor, indexing

INDONESIAN
JOURNAL OF
FUNDAMENTAL
SCIENCES
(IJFS)

E-ISSN: 2621-6728

P-ISSN: 2621-671X

Submitted: February, 19th 2020

Accepted: March, 02nd 2020



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/)

PENDAHULUAN

Kulit merupakan organ tubuh yang memiliki peranan penting. Tanpa kulit, maka otot, tulang dan organ-organ tubuh lainnya tidak bisa berhubungan satu sama lain. Kulit juga berfungsi untuk menjaga suhu tubuh dan membantu merasakan sentuhan. Karena begitu pentingnya fungsi kulit, setiap gangguan yang terjadi pada kulit penting untuk ditangani secara cepat dan tepat untuk mencegah terjadinya efek yang semakin buruk pada tubuh manusia. Menilik perkembangan teknologi yang semakin pesat serta ilmu pengetahuan yang semakin berkembang dibutuhkan sebuah teknologi yang dapat digunakan untuk membantu manusia dalam melakukan diagnosa yang memiliki kemampuan seperti seorang dokter untuk melakukan diagnosa secara cepat dan tepat. Kecerdasan buatan adalah salah satu cabang ilmu pengetahuan yang berhubungan dengan pemanfaatan mesin untuk menyelesaikan suatu permasalahan dengan meniru penalaran/pemikiran manusia.

Salah satu bentuk metode kecerdasan buatan adalah *Case Based Reasoning* (CBR). CBR adalah metode penalaran berbasis kasus yang menggunakan pengetahuan lama atau kumpulan kasus-kasus lama untuk menyelesaikan masalah baru (Russell & Norvig, 2002). Kasus-kasus yang tersimpan dalam rekam medik mengenai diagnosis penyakit kulit serta proses penanganannya dapat digunakan kembali untuk mendiagnosis jenis penyakit kulit yang diderita oleh pasien baru sebagai permasalahan baru dengan menggunakan metode *case based reasoning* (CBR). CBR mengumpulkan kasus sebelumnya yang hampir sama dengan masalah yang baru dan berusaha untuk memodifikasi solusi agar sesuai dengan kasus yang baru (Kolodner, 2014). Dalam CBR untuk mendapatkan dan menemukan solusi terhadap masalah baru digunakan proses pencarian nilai similaritas antara kasus lama dengan masalah baru.

Terdapat dua proses dalam proses *retrieval* yaitu proses *indexing* dan pencarian *similarity*, dimana proses *indexing* merupakan proses yang bersifat *optional* yakni CBR dapat dijalankan tanpa proses *indexing*. Oleh sebab itu dalam penelitian ini peneliti bermaksud melakukan perbandingan hasil akurasi diagnosa dan kecepatan dalam melakukan proses *retrieval* pada CBR menggunakan *indexing* SOM dan tanpa menggunakan *indexing*.

Penggunaan kombinasi CBR dan SOM sebagai metode *indexing* dalam sistem rekomendasi musik berdasarkan konteks. Hasil nilai *precision* sistem dengan menggunakan SOM sebagai metode *indexing* dalam menentukan rekomendasi musik sebesar 0,707 artinya kesesuaian antara mood pengguna dan musik yang direkomendasikan sebesar 70,7% (GIRI & Harjoko, 2015). (Papuangan, 2019) mengimplementasikan metode *nearest neighbor* untuk diagnosis penyakit hepatitis dengan total kasus uji 35 data kasus dan 82 kasus yang disimpan sebagai basis kasus. Hasil pengujian menunjukkan bahwa sistem mampu mengenali tiga jenis penyakit hepatitis dengan tingkat akurasi sebesar 94,29%. (SANTOSO, 2017) membandingkan metode *indexing* SOM dan DBSCAN pada CBR menggunakan dua data yaitu data penyakit gizi buruk pada balita dan penyakit jantung. Membandingkan tiga metode pencarian similaritas yaitu *nearest neighbor*, *euclidan*

distance, dan *minkowski distance*. Hasil penelitian menghasilkan akurasi 100% pada *threshold* ≥ 90 untuk data penyakit gizi buruk pada balita dan data penyakit jantung, didapat dengan menggunakan metode *minkowski distance* diimplementasikan pada CBR dengan *indexing* SOM dan DBSCAN. (Wahyudi & Hartati, 2017) mengimplementasikan CBR untuk diagnosis penyakit Jantung dengan menggunakan metode *nearest neighbor similarity*, *minkowski distance similarity*, dan *euclidean distance similarity*.

Hasil pengujian terhadap data uji penyakit jantung menunjukkan bahwa dengan nilai *threshold* sebesar 80, sistem memiliki unjuk kerja dengan tingkat akurasi menggunakan metode *nearest neighbor similarity* sebesar 86,21%, metode *minkowski distance similarity* sebesar 100% dan metode *euclidean distance similarity* sebesar 94,83%. (Tursina, 2014) mengimplementasikan CBR untuk menentukan daerah berpotensi demam berdarah. Dengan menggunakan metode similaritas *Simple Matching Coefficient* dengan nilai *threshold* 0,95. Dari 20 kasus yang diuji terdapat 1 kasus yang berbeda, tingkat akurasi yang dihasilkan sebesar 95%. Guessoum, Laskri, & Lieber (2014) mendiagnosis penyakit paru menahun dengan fokus pada *missing value* teknik retrieval *nearest neighbor*, keduanya sama-sama menggunakan *rule-based adaptation*.

METODE PENELITIAN

Deskripsi sistem

Sistem yang akan dibangun adalah sistem cerdas berbasis *web* yang mampu melakukan diagnosa penyakit kulit pada manusia menggunakan sistem *Case Based Reasoning* (CBR). Proses pencarian solusi dilakukan dengan dua skenario yaitu skenario pertama diawali dengan pencarian *index* dengan *Self Organizing map*, kemudian melakukan perhitungan similaritas masalah baru dengan kasus-kasus lama (basis kasus). Metode similaritas yang digunakan adalah *Nearest Neighbor*. Skenario kedua adalah menghitung similaritas kasus lama dengan masalah baru tanpa proses *indexing* menggunakan metode similaritas *Nearest Neighbor*.

Proses revisi oleh pakar akan dilakukan apabila masalah baru yang didiagnosa mempunyai nilai dibawah *threshold* atau tidak berhasil didiagnosa (similaritas sama dengan 0). Nilai *threshold* yang digunakan dalam penelitian ini untuk mencari solusi masalah baru adalah ≥ 80 . Nilai tersebut juga digunakan sebagai indikator apakah masalah yang baru tersebut akan di-*retain* ke basis kasus atau tidak. Kasus yang telah direvisi oleh pakar akan tersimpan di basis kasus, secara langsung sistem akan mendapatkan pengetahuan baru dari masalah baru yang telah di revisi. Sedangkan untuk masalah yang memiliki nilai similaritas sama dengan satu yang artinya sama persis dengan kasus sebelumnya tidak disimpan kedalam basis kasus.

Representasi kasus

Data yang digunakan adalah dataset hasil publikasi dari *University of California at Irvine* (UCI) (Murphy, 1994) maka tingkat keyakinan masing-masing kasus adalah 100% dan bobot masing-masing fitur adalah bernilai 1 (satu). Untuk fitur sejarah keturunan (G011) memiliki nilai 1 jika ada dalam keluarga penyakit ini diamati, dan 0 sebaliknya. Fitur usia (G034) memiliki nilai *linear*. Sedangkan fitur lainnya (atribut klinis dan atribut penampakan sel) diberikan nilai 0 sampai 3. 0 mengindikasikan bahwa gejala/fitur tidak dirasakan, 1 mengindikasikan gejala tersebut dirasakan namun hanya sedikit, 2 mengindikasikan gejala tersebut cukup dirasakan, dan 3 mengindikasikan gejala tersebut sangat dirasakan.

Tabel 1. Solusi Kasus

Kode Kasus	Usia	Penyakit
K001	30	P002

Data yang telah diperoleh dari proses akuisisi pengetahuan di representasikan ke dalam bentuk representasi kasus model *relational representation*. Setiap objek atau kasus direpresentasikan oleh sebuah baris dalam sebuah tabel relasional dimana kolom digunakan untuk mendefinisikan atribut atau field dari objek. Sehingga sebuah tabel dapat dibagi menjadi dua bagian yaitu bagian deskripsi masalah dan bagian solusi (Prajarini, 2016).

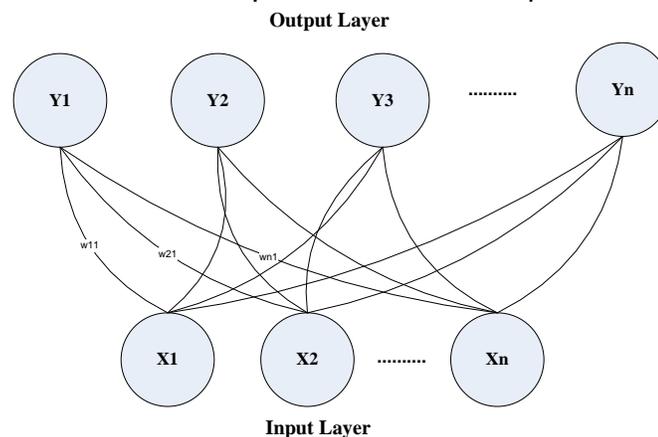
Tabel 2. Detail Kasus

Kode Kasus	Kode Gejala	Nama Gejala	Nilai Derajat Gejala
K001	<i>Clinical attributes</i>		
	G034	Age	54
	G001	Erythema	2
	G002	Scaling	2
	G004	Itching	3
	<i>Histopatological attributes</i>		
	G009	knee and elbow involvement	1
	G016	Exocytosis	3
	G017	Acanthosis	2
	G028	Spongiosis	3
	G032	inflammatory monoluclear infiltrate	1

Indexing

Indexing adalah proses pengelompokan kasus berdasarkan fitur yang ditentukan. *Indexing* berguna untuk mengefisienkan waktu dan memori pada saat proses *retrieve*. Sistem tidak perlu melakukan pencarian nilai kemiripan masalah baru terhadap semua kasus yang ada dalam basis kasus. Penghitungan nilai kemiripan kasus cukup dilakukan terhadap kasus yang ada pada kelompok yang sama.

Metode *indexing* yang digunakan adalah *self organizing maps* (SOM). Arsitektur jaringan SOM tidak memiliki lapis antara (lapisan tersembunyi/*hidden layer*) diantara neuron input dan output. Setiap neuron dalam lapisan input terhubung dengan setiap neuron pada lapisan output. Antar lapisannya dihubungkan dengan bobot tertentu yang dinamakan sebagai vektor bobot (*weight vector*) (Deboeck & Kohonen, 2013). Bobot berfungsi sebagai penentu kedekatan vektor dengan vektor masukan yang diberikan. Setiap neuron dalam lapisan output merepresentasikan kelas dari input yang diberikan. Selama proses penyusunan diri, *cluster* yang memiliki vektor bobot paling cocok dengan pola input (memiliki jarak paling dekat) akan terpilih sebagai pemenang (*winner vector* atau *best matching unit*). Ukuran kedekatan yang sering dipakai adalah jarak *Euclidean*. Vektor pemenang beserta vektor-vektor disekitarnya akan dimodifikasi bobotnya. Kondisi penghentian iterasi adalah ketika mencapai nilai maksimum *epoch*.



Gambar 1 Arsitektur jaringan *Self Organizing Map*

Algoritme *Self Organizing Map* (Deboeck & Kohonen, 2013):

Step 0. Inisialisasi awal bobot w_{ij} (*random*), Inisialisasi nilai parameter *learning rate* (α) dan radius tetangga (R).

Step 1. Selama kondisi bernilai salah, lakukan langkah 2-8

Step 2. Untuk setiap input vektor x , (x_i , $i = 1, 2, \dots, n$) lakukan langkah 3 – 5

Step 3. Hitung jarak *Euclidean Distance* untuk semua j ($j = 1, 2, \dots, m$)

$$D_{(j)} = \sum_i (w_{ij} - x_i)^2 \quad (1)$$

Step 4. Tentukan indeks sedemikian sehingga $D_{(j)}$ minimum

Step 5. Melakukan perbaikan W_{ij} dengan nilai tertentu, yaitu :

$$W_{ij}(\text{baru}) = W_{ij}(\text{lama}) + \alpha [X_i - W_{ij}(\text{lama})] \quad (2)$$

Step 6. Perbaiki *learning rate*.

Step 7. Kurangi radius ketetanggaan topologi dengan pencacahan tertentu.

Step 8. Uji kondisi berhenti, jika benar maka kondisi berhenti.

Kondisi penghentian iterasi pada penelitian ini adalah ketika telah mencapai maksimum iterasi maka proses dapat dihentikan.

Retrieval dan similarity

Similaritas dalam *Case Based Reasoning* adalah tingkat kemiripan suatu kasus lama dengan masalah baru. Dalam proses *retrieval* kasus-kasus lama dibandingkan dengan masalah baru kemudian dihitung nilai kemiripannya. Fungsi similaritas menghasilkan nilai yang akan menentukan tentang ada atau tidak adanya kemiripan antara kasus lama yang tersimpan dalam basis kasus dengan permasalahan baru. Pencarian kemiripan antara kasus lama dan masalah baru tergantung pada cara mengukur keserupaan antar fitur. Nilai similaritas antara 0 hingga 1, 0 berarti tidak ada kemiripan antara kasus lama dengan kasus baru sedangkan 1 adalah ditemukan kemiripan antara kasus lama dengan kasus baru yang berarti kemiripan sebesar 100% sama. Metode similaritas yang digunakan pada penelitian ini adalah *Nearest Neighbor (SimNN(T,S))* yang telah dimodifikasi menggunakan keyakinan kasus (Mancasari, 2012).

$$SimNN(T,S) = \frac{\sum_{i=1}^n f_i(S_i,T_i)(W_{i,p(s)})}{\sum_{i=1}^n (W_{i,p(s)})} * P(S) * \frac{J(S_i,T_i)}{J(T_i)}$$

dimana $f(S_i,T_i)$ adalah kesamaan fitur ke- i dari kasus S (*source*) dan kasus T (*target*), S_i merupakan nilai fitur ke- i dari kasus lama (*source*), T_i menunjukkan nilai fitur ke- i dari masalah baru (*target*). " w_i " adalah bobot untuk fitur ke- i , " $P(S)$ " merupakan persentase tingkat keyakinan pakar terhadap suatu kasus dalam *source case*. " $J(S_i,T_i)$ " adalah banyaknya fitur yang terdapat dalam *target case* yang muncul pada fitur *source case* sedangkan " $J(T_i)$ " adalah banyaknya fitur yang terdapat dalam *target case*.

HASIL DAN PEMBAHASAN

Proses pelatihan basis kasus

Proses pelatihan basis kasus menggunakan algoritme *self organizing map* dilakukan untuk memperoleh nilai pusat cluster yang akan digunakan pada proses *indexing* masalah baru. Data gejala sebagai *input* pelatihan pada proses pelatihan. Parameter yang digunakan pada proses pelatihan basis kasus yaitu nilai *learning rate*, maksimum *epoch*, jumlah *cluster*, dan penurunan pembelajaran. Proses penentuan nilai parameter dilakukan dengan menguji beberapa nilai untuk masing-masing parameter. Pengujian parameter *learning rate* dengan nilai 0,1 – 0,9 dengan nilai *increment* 2. Pengujian parameter maksimum *epoch* dengan nilai 50, 100, dan 150.

Tabel 3. Pengujian kondisi perhentian proses pelatihan SOM

Kondisi Berhenti	Akurasi proses retrieve
Epoch = 50	92,52%
Epoch = 100	92,52%
Epoch = 150	92,52%

Pada table diatas terlihat pada epoh 50 hingga 150 memiliki akurasi yang sama yakni 92,52% maka pada penelitian ini menggunakan epoh paling minimal yakni 50. Setelah dilakukan pengujian parameter epoh, selanjutnya dilakukan pengujian parameter *learning rate* menggunakan epoh 50.

Tabel 4. Hasil pengujian parameter *learning rate*

Learning Rate	0,1	0,3	0,5	0,7	0,9
Akurasi	92,52%	92,52%	93,46%	92,52%	92,52%

Dari tabel diatas tampak, pada *learning rate* 0,5 memiliki akurasi lebih tinggi dibandingkan nilai *learning rate* lainnya, yaitu sebesar 93,46%. Maka berdasarkan dua pengujian parameter tersebut pada penelitian ini menggunakan nilai epoh 50 dan *learning rate* 0,5. Nilai parameter ini tidak berubah dan digunakan dalam proses retain.

Hasil pengujian

Pengujian sistem CBR menggunakan *indexing* SOM metode *nearest neighbor* pada *threshold* ≥ 70 memiliki akurasi yang sama dengan pengujian CBR tanpa metode *indexing*, begitu pula pada *threshold* ≥ 90 memberikan hasil yang sama masing-masing sebesar 68,22%.

Tabel 5. Perbandingan akurasi CBR dengan *indexing* SOM dan tanpa *indexing*

Metode Similaritas	CBR tanpa Indexing			Rata-rata waktu retrieve
	Threshold			
	$\geq 70\%$	$\geq 80\%$	$\geq 90\%$	
Nearest Neighbor	95,33%	93,46%	68,22%	0,469
Metode Similaritas	CBR dengan Indexing SOM			Rata-rata waktu retrieve
	Threshold			
	$\geq 70\%$	$\geq 80\%$	$\geq 90\%$	
Nearest Neighbor	95,33%	92,52%	68,22%	0,15463

Pada *threshold* ≥ 80 menunjukkan hasil yang berbeda yakni CBR tanpa *indexing* memberikan hasil akurasi sebesar 93,46% sedangkan CBR dengan *indexing* SOM memberikan hasil akurasi sebesar 92,52%. Hal ini disebabkan karena pada CBR tanpa *indexing* sistem melakukan proses *retrieval* atau pengujian disemua basis kasus yang ada dalam basisdata sistem sehingga untuk kasus-kasus lama yang memiliki kemiripan tinggi dengan kasus baru pasti akan ikut dalam proses *retrieval*. Sedangkan jika menggunakan *indexing* hanya beberapa kasus yang dilakukan perbandingan similaritas antara kasus lama dengan kasus baru, sehingga lingkup pencarian kasusnya yang serupa jauh lebih kecil. Namun, kelebihan dari CBR menggunakan *indexing* berdasarkan pengujian kecepatan waktu pencarian similaritas kasus, CBR dengan *indexing* jauh lebih cepat memberikan hasil diagnose dibandingkan dengan CBR tanpa metode *indexing*.

KESIMPULAN

1. Proses diagnosis yang memiliki akurasi paling baik adalah CBR tanpa menggunakan metode *indexing* SOM dengan hasil akurasi 93,46% pada *threshold* ≥ 80 .
2. Kecepatan waktu pada proses *retrieval* CBR dengan *indexing* SOM relative lebih cepat dibandingkan dengan tanpa menggunakan *indexing*. Rata-rata waktu *retrieve* tanpa *indexing* adalah 0,469 detik sedangkan menggunakan *indexing* SOM hanya membutuhkan waktu 0,15 detik. Hal ini dikarenakan dalam proses *retrieve* sistem hanya mencari nilai kemiripan pada kasus yang masuk ke dalam *cluster* yang sesuai dengan masalah baru, sehingga proses diagnosis menjadi lebih cepat.

DAFTAR PUSTAKA

- Deboeck, G., & Kohonen, T. (2013). *Visual explorations in finance: with self-organizing maps*. Springer Science & Business Media.
- GIRI, G. S. T. A. Y. U. V. M., & Harjoko, A. (2015). *Sistem Rekomendasi Musik Berdasarkan Konteks Menggunakan Case-Based Reasoning dan Self Organizing Map*. Universitas Gadjah Mada.
- Guessoum, S., Laskri, M. T., & Lieber, J. (2014). RespiDiag: a case-based reasoning system for the diagnosis of chronic obstructive pulmonary disease. *Expert Systems with Applications*, 41(2), 267–273.
- Kolodner, J. (2014). *Case-based reasoning*. Morgan Kaufmann.
- Mancasari, U. K. (2012). *Sistem Pakar Menggunakan Penalaran Berbasis Kasus Untuk Mendiagnosa Penyakit Saraf Pada Anak*. Skripsi. Prodi S1 Ilmu Komputer, Universitas Gadjah Mada, Yogyakarta.
- Murphy, P. M. (1994). UCI repository of machine learning databases. *Ftp:/Pub/Machine-Learning-Databaseonics. Uci. Edu*.
- Papuangan, M. (2019). Penerapan Case Based Reasoning Untuk Sistem Diagnosis Penyakit Hepatitis. *JIKO (Jurnal Informatika Dan Komputer)*, 1(1), 7–12.
- Prajarini, D. (2016). Perbandingan Algoritma Klasifikasi Data Mining Untuk Prediksi Penyakit

- Kulit. *INFORMAL: Informatics Journal*, 1(3), 137–141.
- Russell, S., & Norvig, P. (2002). *Artificial intelligence: a modern approach*.
- SANTOSO, H. (2017). *Metode Indexing pada Case Based Reasoning (CBR) menggunakan Density Based Spatial Clustering Application with Noise (DBSCAN)*. Universitas Gadjah Mada.
- Tursina, T. (2014). *Case Base Reasoning untuk Menentukan Daerah Menentukan Daerah Berpotensi Demam Berdarah (Studi Kasus Kota Pontianak)*. *Seminar Nasional Bisnis Dan Teknologi 2014*. Darmajaya Informatics and Business Institute.
- Wahyudi, E., & Hartati, S. (2017). *Case-Based Reasoning untuk Diagnosis Penyakit Jantung*. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 11(1), 1–10.