

PENALARAN BERBASIS KASUS UNTUK MENDIAGNOSA PENYAKIT INFEKSI MENULAR SEKSUAL (IMS) MENGGUNAKAN ALGORITMA *WEIGHTED EUCLIDEAN DISTANCE*

Derin N. Liu¹, Sebastianus A. S. Mola² dan Yelly Y. Nabuasa³

^{1,2,3} Program Studi Ilmu Komputer, Fakultas Sains dan Teknik, Universitas Nusa Cendana

¹Email: derinnoraeniliu@gmail.com

²Email: adimola@staf.undana.ac.id

³Email: yelly.yosiana.n@gmail.com

ABSTRAK

Penalaran Berbasis Kasus adalah sebuah metodologi untuk penyelesaian masalah dengan memanfaatkan pengalaman sebelumnya. Pada penelitian ini penulis menerapkan penalaran berbasis kasus untuk mendiagnosa penyakit infeksi menular seksual menggunakan metode *weighted euclidean distance*. Sumber basis pengetahuan diperoleh dengan mengumpulkan berkas rekam medis pasien penyakit infeksi menular seksual pada tahun 2016-2017. Proses pencarian solusi dimulai dengan mengeliminasi data yang tidak relevan menggunakan C4.5 dan berlanjut dengan perhitungan nilai kemiripan menggunakan algoritma *weighted euclidean distance*. Sistem ini dapat mendiagnosis 5 jenis penyakit infeksi menular seksual berdasarkan 123 gejala yang ada. Hasil sistem berupa jenis penyakit infeksi menular seksual berdasarkan gejala yang dialami pasien, solusi pengobatan dan presentasi kemiripan kasus baru dengan kasus lama. Berdasarkan hasil pengujian dengan 127 kasus infeksi menular seksual (IMS) didapatkan hasil: Pengujian menggunakan skenario *K-Fold Cross Validation*, total data dibagi menjadi 10 *fold* dan proses pengujian dibagi menjadi 2 bagian yaitu pengujian menggunakan *indexing* dan pengujian tanpa menggunakan *indexing*. Untuk pengujian menggunakan *indexing* akurasi tertinggi yang didapat sebesar 90.84% pada *fold* ke-2, dan rata-rata akurasi dari keseluruhan *fold* adalah sebesar 88.55% dengan rata-rata waktu yang dihasilkan 9498 ms (milidetik) sedangkan pengujian tanpa menggunakan *indexing* akurasi tertinggi yang didapat sebesar 63.03% pada *fold* ke-2, dan rata-rata akurasi dari keseluruhan *fold* adalah sebesar 53.48% dengan rata-rata waktu yang dihasilkan 9975 ms (milidetik).

Kata kunci: Penalaran berbasis kasus, Algoritma C4.5, Algoritma *Weighted Euclidean Distance*, Penyakit Infeksi Menular Seksual

ABSTRACT

Case-based reasoning is a methodology for solving problems by utilizing previous experience. In this study the authors apply case-based reasoning to diagnose sexually transmitted infection using the weighted Euclidean distance method. Source of the knowledge base was obtained by collecting medical record of patients with sexually transmitted infections in 2016-2017. The process of finding a solution starts with eliminating irrelevant data using the C4.5 method and continues with the calculation of the similarity value using the Weighted Euclidean Distance algorithm. This system can diagnose 5 types of sexually transmitted infections based on 123 existing symptoms. System result in the form of sexually transmitted infections based on symptoms experienced by the patient, treatment solution and presentation of similarities between new cases and old cases. Based on the result of testing with 127 cases of sexually transmitted infections obtained result: testing uses the K-Fold Cross Validation scenario, the total data is divided into 10fold and the testing process is divided into 2 parts, namely testing using indexing and testing without using indexing. For testing using the highest accuracy indexing obtained at 90.84% in the second fold, and the average accuracy of the entire fold is 88.55% with the average time generated 9498 ms (millisecond), while testing without using the highest accuracy indexing obtained by 63.03% in the second fold, and the average accuracy of the entire fold is 53.48% with the average time generated 9975 ms (millisecond).

Keywords: Case-based reasoning, C4.5 Algorithm, Weighted Euclidean Distance Algorithm, Sexually Transmitted Infectious Diseases.

1. PENDAHULUAN

Kesehatan merupakan sebuah kebutuhan yang mendasar bagi setiap orang. Menurut H.L. Bloom menyatakan bahwa ada empat faktor yang mempengaruhi derajat kesehatan secara berturut-turut, yaitu: 1)

gaya hidup (*life style*), 2) lingkungan (sosial, ekonomi, politik, budaya) 3) pelayanan kesehatan dan 4) faktor genetik (keturunan). Dalam kehidupan sehari-hari ada beberapa penyakit yang mempengaruhi kesehatan setiap individu dalam lingkungan diantaranya, penyakit menular dan penyakit tidak menular. Infeksi menular seksual merupakan salah satu penyakit yang bisa ditularkan. IMS adalah penyakit yang disebarkan oleh hubungan seks. Di Indonesia sendiri, telah banyak laporan mengenai prevalensi infeksi menular seksual dari beberapa lokasi antara tahun 1999 sampai 2001 menunjukkan prevalensi infeksi gonore dan klamidia yang tinggi antara 20%-35% [1].

Pada penelitian sebelumnya yang berjudul “Perancangan Sistem Pakar Diagnosa Infeksi Menular Seksual (IMS) Berbasis Web dengan Metode Forward dan Backward Chaining pada RSUP Dr. Sardjito Yogyakarta”. Web sistem pakar ini menggunakan 4 faktor resiko, 3 keluhan, dan 65 gejala dan 12 jenis penyakit dengan baik, yaitu sudah mendekati hasil diagnosis yang dibuat seorang dokter penyakit infeksi menular seksual dengan presentasi 75%.

Permasalahan yang sering muncul adalah tidak semua penderita sakit (pasien) memiliki keberanian yang cukup untuk datang ke dokter dan melakukan pemeriksaan dikarenakan malu apabila penyakitnya diketahui oleh orang lain sehingga tidak sedikit orang yang juga akhirnya memutuskan untuk mengikuti kebiasaan dari nenek moyang yang lebih cenderung mempercayai khasiat dari obat tradisional di bandingkan obat dari dokter padahal penyakit tersebut cukup serius sehingga dibutuhkan penanganan khusus dari dokter/ahli agar bisa memberikan solusi yang tepat berdasarkan penyakit yang diderita oleh pasien.

Cara kerja *case based reasoning* (CBR) adalah dengan membandingkan kasus baru dengan kasus lama, jika kasus baru tersebut mempunyai kemiripan dengan kasus lama maka CBR akan memberikan jawaban kasus lama untuk kasus baru tersebut. Jika tidak ada yang cocok maka CBR akan melakukan adaptasi dengan cara memasukkan kasus baru kedalam database penyimpanan kasus (*case based*), sehingga tidak secara langsung pengetahuan CBR akan bertambah. Itu adalah kelebihan utama dari CBR dibanding *rule based*.

2. MATERI DAN METODE

Data Penelitian

Data rekam medis yang dipakai dalam penelitian ini diperoleh dari Puskesmas Oepoi Kupang pada tahun 2016-2017 berjumlah 127 kasus.

Penalaran berbasis kasus

Penalaran berbasis kasus/*case based reasoning* adalah sebuah metodologi untuk penyelesaian masalah dengan memanfaatkan pengalaman sebelumnya [2]. Di dalam CBR, seseorang yang melakukan penalaran dapat menyelesaikan masalah baru dengan memperhatikan kesamaannya dengan satu atau beberapa penyelesaian dan permasalahan sebelumnya. Salah satu tahap penting dalam siklus CBR adalah pengambilan kembali (*retrieval*) terhadap kasus-kasus sebelumnya yang dapat digunakan untuk menyelesaikan masalah baru. Proses *reuse* dalam siklus CBR adalah tanggung jawab dalam memberikan solusi dari sebuah kasus baru berdasarkan penyelesaian-penyelesaian kasus yang diambil kembali. Dalam kondisi ini, penyelesaian yang diambil kembali memerlukan adaptasi untuk mengatasi perbedaan-perbedaan penting tersebut. Pembuatan keputusan medis merupakan salah satu bidang yang memerlukan adaptasi. *Retain* merupakan tahap terakhir dalam siklus CBR yang menghasilkan penyelesaian masalah terbaru yang digabung dalam sistem pengetahuan. Hal ini telah diterjemahkan menjadi berbagai pendekatan untuk merekam hasil dari penyelesaian masalah sebagai sebuah kasus baru dan dapat ditambahkan dalam basis kasus.

Indexing

Indexing merupakan proses pengelompokkan kasus yang ada berdasarkan fitur yang ditentukan. Proses *indexing* dapat mengefisienkan waktu dan memory karena pada saat melakukan proses pencarian kasus yang mirip dengan kasus yang baru, sistem CBR tidak perlu menghitung nilai kemiripan terhadap kasus yang berada pada kelompok yang sama dengan kasus baru tersebut.

Algoritma C4.5

Algoritma C4.5 adalah algoritma yang digunakan untuk menghasilkan sebuah pohon keputusan yang dikembangkan oleh Ross Quinlan. Ide dasar dari algoritma ini adalah pembuatan pohon keputusan berdasarkan pemilihan atribut yang memiliki prioritas tertinggi atau dapat disebut memiliki nilai *gain*

tertinggi berdasarkan nilai *entropy* atribut tersebut sebagai poros atribut klasifikasi. Rumus perhitungan nilai *entropy* dan *gain* ditunjukkan pada persamaan 1 dan 2.

$$Entropy (S) = \sum_{i=1}^n - p_i * \log_2 p_i \dots\dots\dots (1)$$

di mana:

S: Himpunan kasus

n: Jumlah partisi

p_i : Jumlah kasus pada partisi ke- i

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy (S_i) \dots\dots\dots (2)$$

di mana:

S: Himpunan kasus

N: Jumlah partisi atribut A

$|S_i|$: Jumlah kasus pada partisi ke- i

$|S|$: Jumlah kasus dalam s

Weighted Euclidean Distance

Perhitungan untuk mendapatkan kesamaan antar kasus, sangat penting pada metode CBR terutama dalam proses *retrieval*. Efektifitas dari pengukuran kesamaan tergantung dari seberapa banyak kasus yang ditelaah, membantu dalam memecahkan masalah pada kasus baru [3]. Rumus perhitungan *Weighted Euclidean Distance* ditunjukkan pada persamaan 3.

$$d(x, y) = \sqrt{\sum_{i=1}^n (W_i * d_i (C_x, C_y))^2} \dots\dots\dots (3)$$

di mana:

$d(x,y)$: perhitungan jarak antara kasus lama dengan kasus baru

W_i : bobot kriteria ke- i

$d_i(C_x, C_y)$: jarak menghitung *Euclidean distance*

C_x : kasus baru

C_y : kasus lama

n: jumlah kriteria

Similarity

Metode *similarity* merupakan tranformasi nilai kedekatan berdasarkan kemiripan dari kasus lama dengan kasus baru. Kasus lama dengan kasus baru dibandingkan antara atribut dengan atributnya. Kemudian nilai yang diperoleh bernilai 1 apabila kasus lama tersebut mirip dan bernilai 0 apabila kasus lama tersebut tidak mirip [4]. Rumus perhitungan *Weighted Euclidean Distance* ditunjukkan pada persamaan 4.

$$SIM (C_x, C_y) = \left(\frac{1}{1 + d(C_x, C_y)} \right) \dots\dots\dots (4)$$

di mana:

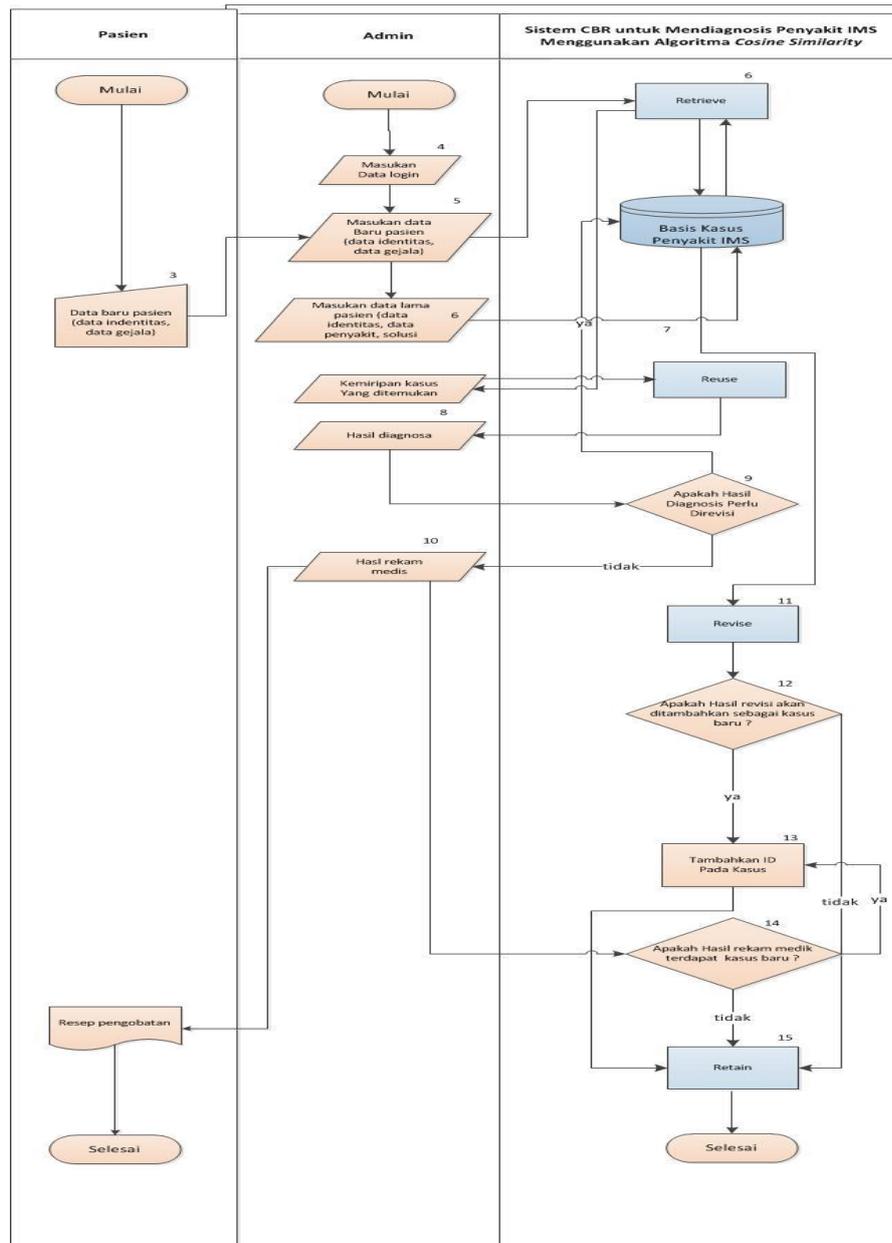
$SIM (C_x, C_y)$: nilai similarity dari kasus lama dengan kasus baru

$d(C_x, C_y)$: hasil perhitungan jarak dengan metode *Weighted Euclidean Distance*

Tahapan perancangan sistem CBR

Gambar 1 menjelaskan *flowchart* sistem CBR ini yang dimulai dengan entitas pasien memberikan data diri dan data gejala kepada admin. Admin melakukan *login* dan memasukan data baru kasus pasien (data identitas, data gejala) ke sistem CBR untuk diolah. Selanjutnya sistem akan memberikan hasil diagnosis yang tepat berdasarkan kesamaan dengan kasus lama pada basis kasus.

Jika informasi mengenai hasil diagnosis tidak direvisi berarti kasus tersebut memiliki nilai kemiripan lebih besar atau sama dengan nilai *threshold* dengan kasus lama yang terdapat dalam basis kasus dan sistem langsung mengeluarkan informasi mengenai hasil rekam medis berdasarkan hasil diagnosis sehingga pasien akan menerima resep pengobatan dari admin.



Gambar 1. Flowchart sistem CBR diagnosis penyakit IMS

3. HASIL DAN PEMBAHASAN

Pengujian sistem yang bertujuan untuk mengukur keakuratan sistem dalam mendiagnosa penyakit infeksi menular seksual dilakukan menggunakan pengujian *K-Fold Validation*. Pengujian sistem ini dilakukan dengan membagi 127 data kedalam 10 *fold* yang kemudian dibagi menjadi dua bagian yaitu data latih dan data uji. Proses perhitungan yang dilakukan yaitu menggunakan *indexing* dan tanpa menggunakan *indexing*. Hasil akurasi menggunakan *indexing* tertinggi yang didapatkan sebesar 90.84% pada *fold* ke-2, dan rata-rata nilai hasil akurasi dari keseluruhan *fold* adalah 88.55% dengan rata-rata waktu yang dihasilkan 9498 ms (milidetik), sedangkan hasil akurasi tanpa *indexing* tertinggi yang didapatkan sebesar 63.03% pada *fold* ke-2, dan rata-rata nilai hasil akurasi dari keseluruhan *fold* adalah 53.48% dengan rata-rata waktu yang dihasilkan 9975 ms (milidetik). Gambaran secara umum dapat dilihat pada tabel 1 dan 2.

Tabel 1. Rata-rata Akurasi Perhitungan *K-Fold* Menggunakan *Indexing*

<i>K-Fold</i> Ke-	Rata-rata akurasi Perhitungan <i>K-Fold</i>
1	88.35%
2	90.84%
3	86.73%
4	88.59%
5	90.56%
6	87%
7	88.13%
8	88.73%
9	89.16%
10	87.44%
Rata-rata Akurasi	88.55%
Rata-rata Waktu Pengujian	9498 ms (milidetik)

Tabel 2. Rata-rata Akurasi Perhitungan *K-Fold* Tanpa Menggunakan *Indexing*

<i>K-Fold</i> Ke-	Rata-rata akurasi Perhitungan <i>K-Fold</i>
1	51.87%
2	63.03%
3	47.65%
4	55.05%
5	60.88%
6	47.65%
7	51.21%
8	52.74%
9	56.19%
10	48.57%
Rata-rata Akurasi	53.48%
Rata-rata Waktu Pengujian	9975 ms (milidetik)

4. KESIMPULAN DAN SARAN

Kesimpulan

Sistem yang dibangun dengan metode *indexing* C4.5 dan *Weighted Euclidean Distance* dapat berjalan dengan baik dan hasil akurasi menggunakan *indexing* tertinggi yang didapatkan sebesar 90.84% pada *fold* ke-2, dan rata-rata nilai hasil akurasi dari keseluruhan *fold* adalah 88.55% dengan rata-rata waktu yang dihasilkan 9498 ms (milidetik), sedangkan hasil akurasi tanpa *indexing* tertinggi yang didapatkan sebesar 63.03% pada *fold* ke-2, dan rata-rata nilai hasil akurasi dari keseluruhan *fold* adalah 53.48% dengan rata-rata waktu yang dihasilkan 9975 ms (milidetik) serta sistem mampu menerapkan semua tahapan CBR (*Retrieve, Reuse, Revise, Retain*) untuk 5 jenis penyakit Keputihan (*flour albus*), Kencing Nanah (*gonore*), Kutil Kelamin (*kondiloma*), Herpes Kelamin, Raja Singa (*sifilis*), dengan 123 gejala.

Saran

Proses pembobotan dalam penelitian ini adalah pembobotan yang dilakukan oleh satu pakar dan tentunya bersifat subjektif, sehingga untuk penelitian selanjutnya disarankan pembobotan dilakukan oleh beberapa pakar dan diambil rata-rata agar diperoleh pembobotan yang lebih baik.

DAFTAR PUSTAKA

- [1] A. A. Wirakusuma, I. G. K. Darmada, and L. M. M. Rusyati, 'Spektrum Infeksi Menular Seksual di Poliklinik Kulit dan Kelamin Rumah Sakit Umum Pusat Sanglah Periode 2009-2011', *E-Jurnal Medika Udayana*, 2014.
- [2] J. Main, T. S. Dillon, and S. C. Shiu, 'A tutorial on case based reasoning', in *Soft computing in case based reasoning*, Springer, pp. 1–28, 2001
- [3] M. Greenacre, *Correspondence analysis in practice*. CRC press, 2017.
- [4] M. M. Richter and R. O. Weber, *Case-based reasoning*. Springer, 2016.