

**PERBANDINGAN METODE SIMILARITAS PADA PENALARAN BERBASIS
KASUS MENGGUNAKAN *SIMPLE MATCHING COEFFICIENT* DAN
*SORENSEN-DICE COEFFICIENT***

Murien Nugraheni¹

**¹Prodi Sistem dan Teknologi Informasi, Fakultas Teknik, Universitas Negeri Jakarta
muriennugraheni@unj.ac.id**

Abstrak

Terdapat pendekatan baru yang dapat digunakan untuk mendiagnosa penyakit, yaitu menggunakan Penalaran Berbasis Kasus. Penalaran berbasis kasus merupakan salah satu teknik untuk pemecahan masalah dengan menggunakan kasus lama yang tersimpan di dalam basis kasus yang nantinya akan digunakan untuk memberikan solusi dari kasus baru. Penalaran berbasis kasus dapat digunakan untuk melakukan diagnosa penyakit berdasarkan gejala-gejala fisik yang dimiliki oleh hewan ternak. Penelitian ini membandingkan metode similaritas menggunakan *Simple Matching Coefficient* dan *Sorensen-Dice Coefficient*. Dari hasil perbandingan kedua metode tersebut, *Simple Matching Coefficient* mempunyai tingkat persentase kemiripan sebesar 46.3422% dan *Sorensen-Dice Coefficient* mempunyai tingkat persentase kemiripan sebesar 53.6577%. Sehingga *Sorensen-Dice Coefficient* mempunyai persentase kemiripan lebih tinggi dibandingkan dengan persentase kemiripan pada *Simple Matching Coefficient*.

Kata kunci : Penalaran Berbasis Kasus, *Simple Matching Coefficient*, *Sorensen-Dice Coefficient*.

1. Pendahuluan

Kolaborasi disiplin ilmu komputer dengan disiplin ilmu lain sudah banyak dilakukan misalnya dengan ilmu kedokteran. Ada beberapa aplikasi komputer yang digunakan untuk membantu kerja dokter sehingga memberikan hasil kerja yang memuaskan, seperti sistem pakar untuk mendiagnosa penyakit. Para peneliti kecerdasan buatan berusaha untuk membuat sistem yang lebih baik lagi sehingga kekurangan-kekurangan yang terdapat pada sistem sebelumnya dapat diperbaiki. Ide dasar dari penalaran berbasis kasus adalah menyelesaikan suatu masalah baru dengan menggunakan jawaban dari masalah lama. Penyajian pengetahuan dibuat dalam bentuk kasus-kasus. Setiap kasus berisi masalah dan jawaban, sehingga kasus lebih mirip dengan suatu pola tertentu. Cara kerja penalaran berbasis kasus adalah dengan membandingkan kasus baru dengan kasus lama, jika kasus baru tersebut mempunyai kemiripan dengan kasus lama maka penalaran berbasis kasus akan memberikan jawaban kasus lama untuk kasus baru tersebut. Jika tidak ada yang cocok maka penalaran berbasis kasus akan melakukan adaptasi dengan memasukkan kasus baru tersebut kedalam basis data penyimpanan kasus, sehingga secara tidak langsung pengetahuan dari penalaran berbasis kasus akan bertambah.

Penelitian yang dilakukan oleh Faizal (2014:83) menerapkan *case-based reasoning* atau penalaran berbasis kasus untuk membangun suatu sistem yang memiliki kemampuan untuk mendiagnosa penyakit *cardiovascular* berdasarkan similaritas pada kasus-kasus terdahulu menggunakan *Simple Matching Coefficient Similarity*. Dari hasil pengujian yang dilakukan dihasilkan bahwa sistem yang dibangun memiliki nilai sensitifitas 97.06%, spesifisitas 64.29%, PPV 86.84%, NPV 90.00%, akurasi 87.50% dengan tingkat kealahan atau *error rate* sebesar 12.50%. Penelitian yang dilakukan oleh Amaliah, dkk (2017:33) menggunakan metode *Case-Based Reasoning*, *Naïve Bayes*, dan *Simple Matching Coefficient* untuk menentukan klasifikasi anak yang mengalami keterlambatan mental. Berdasarkan hasil uji kelayakan sistem yang dilakukan menggunakan kuesioner diperoleh baik, dengan persentase variabel tampilan 48.75%, variabel kemudahan penggunaan 53.33%, variabel kinerja sistem 46.67%, dan variabel isi 51.67%.

Penalaran berbasis kasus mencoba meniru kemampuan manusia dalam menyelesaikan suatu kasus berdasarkan kasus-kasus yang sebelumnya terjadi. Penyelesaian kasus tersebut didasarkan pada pengalaman yang dimiliki ataupun berdasarkan dari pengalaman orang lain. Penyelesaian penalaran berbasis kasus dapat menggunakan metode similaritas untuk menghitung tingkat kemiripan antara dua objek yaitu antara kasus lama yang tersimpan di dalam basis kasus dengan kasus baru. Metode similaritas yang dapat digunakan untuk menghitung tingkat kemiripan yaitu, *Simple Matching Coefficient* dan *Sorensen-Dice Coefficient*.

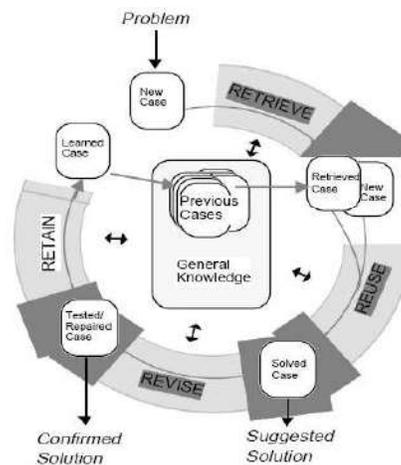
2. Metode

2.1 Penalaran Berbasis Kasus

Siklus pada penalaran berbasis kasus dibagi menjadi 4, antara lain :

- Retrieve* : pencarian kasus lama yang paling mirip dengan kasus baru.
- Reuse* : penggunaan kasus lama untuk memberikan solusi pada kasus baru yang baru saja diinputkan ke dalam basis pengetahuan.
- Revise* : penyesuaian solusi dari kasus lama yang nantinya akan digunakan sebagai solusi pada kasus baru.
- Retain* : penggunaan solusi kasus baru yang kasus baru tersebut diperbaharui atau diinputkan ke dalam basis kasus.

Gambar 1 menunjukkan siklus dari penalaran berbasis kasus, yang terdiri dari *retrieve*, *reuse*, *revise*, dan *retain*.



Gambar 1. Siklus Penalaran Berbasis Kasus (Source : Aamodt dan Plaza,1994)

2.2 *Simple Matching Coefficient*

Penelusuran yang digunakan pada penelitian ini adalah dengan cara membandingkan setiap fitur yaitu antara fitur kasus baru dengan fitur kasus yang ada dibasis kasus, kemudian hasil perbandingan tersebut akan dihitung similaritasnya. Fitur yang digunakan adalah berupa gejala-gejala yang tampak pada hewan ternak.

Simple Matching Coefficient merupakan salah satu untuk menghitung tingkat kemiripan dua buah objek yang bernilai biner. Perbandingan fitur menggunakan biner yaitu 1 untuk menyatakan ada gejala dan 0 untuk menyatakan tidak ada gejala. Perhitungan similaritas digunakan untuk menghasilkan nilai apakah ada kemiripan atau tidak antara kasus baru dengan kasus yang telah ada di basis kasus.

Formula yang digunakan pada *Simple Matching Coefficient* untuk menghitung tingkat kemiripan antara dua objek x dan y adalah sebagai berikut :

$$SMC(x,y) = \frac{M_{11}+M}{M_{10}+M_{01}+M} \text{---(Tursina, 2012)}$$

dimana :

x : kasus lama

y : kasus baru

M_{11} : jumlah atribut biner, x=1 dan y=1

M_{10} : jumlah atribut biner, x=1 dan y=0

M_{01} : jumlah atribut biner, x=0 dan y=1

M_{00} : jumlah atribut biner, x=0 dan y=0

Kasus baru (y) adalah kasus yang akan dicari solusinya dengan cara membandingkan fitur gejala pada setiap kasus lama (x) atau kasus yang tersimpan di basis kasus. Kasus baru dikatakan similar (mirip) 100% dengan kasus yang lama apabila nilai similaritas dari $SMC(x,y)$ sama dengan 1 sedangkan tidak similar apabila nilai $SMC(x,y)$ sama dengan 0. Pada SMC semua jumlah atribut biner (11, 00, 10, 01) dimasukkan dalam perhitungan similaritasnya.

2.3 *Sorensen-Dice Coefficient*

Salah satu cara untuk menghitung similaritas dua objek (items) yang bersifat biner adalah dengan menggunakan metode *Sorensen-Dice Coefficient*. Berikut adalah formula dari *Sorensen-Dice Coefficient*:

$$Sorensen - Dice = \frac{2a}{(2a+b+c)} \text{---(Chao, A., 2006)}$$

3. Hasil dan Pembahasan

3.1 Representasi Kasus

Pada representasi kasus setiap kasus memiliki gejala-gejala yang akan digunakan dalam pencarian kasus yang mirip. Contoh representasi kasus dapat dilihat pada Tabel 1.

Tabel 1 berisi relasi antara gejala penyakit dengan kasus yang terjadi. Kasus-kasus yang sudah dikumpulkan akan direpresentasikan ke dalam bentuk tabel. Seperti yang terlihat pada Tabel 1. Pada tabel tersebut terdapat 8 gejala yaitu G1 hingga G8 dan 20 kasus yaitu K1 hingga K20.

Nilai yang terdapat pada Tabel 1 merupakan nilai yang merepresentasikan gejala yang terjadi pada hewan ternak.

Nilai tersebut terbagi menjadi 2 jenis, yaitu :

- 0 yang berarti gejala tersebut tidak terlihat
- 1 yang berarti gejala tersebut sangat terlihat

Tabel 1. Representasi Kasus Lama dan Kasus Baru

Kasus	G1	G2	G3	G4	G5	G6	G7	G8
K1	1	1	0	1	1	0	1	1
K2	1	0	0	0	0	1	1	0
K3	0	1	1	0	1	0	0	0
K4	0	1	1	1	0	1	1	1
K5	0	1	1	0	0	0	0	0
K6	0	0	1	1	0	0	0	0
K7	0	1	1	0	0	1	1	1
K8	0	0	1	0	0	1	0	0
K9	0	1	1	0	1	0	0	1
K10	1	1	0	0	0	0	0	0
K11	0	1	1	0	0	0	0	0
K12	0	0	1	1	1	0	0	0
K13	1	1	0	0	1	0	0	0
K14	0	1	1	0	0	0	0	1
K15	0	1	0	1	0	0	0	0
K16	0	1	0	0	0	1	1	0
K17	0	0	1	0	0	1	0	0
K18	1	1	1	0	0	0	0	0
K19	0	1	0	1	0	0	0	0
K20	1	1	0	1	1	0	1	0
K_{baru} (y)	1	1	1	1	0	1	1	1

Keterangan :

- 0 yang berarti gejala tersebut **tidak terlihat**
- 1 yang berarti gejala tersebut **sangat terlihat**
- Kasus : K1, K2, K3, K4, K5, ..., K20
- Gejala : G1, G2, G3, G4, G5, G6, G7, G8

G1 : gangguan syaraf

G2 : diare

G3 : nafsu makan berkurang

G4 : merejan

G5 : tubuh gemetar

G6 : bulu kusam/ kotor

G7 : bulu sekitar anus kotor

G8 : lesu

3.2 Perhitungan Tingkat Kemiripan *Simple Matching Coefficient*

Hasil perhitungan similaritas dengan menggunakan *Simple Matching Coefficient* yaitu :

Tabel 2. Perhitungan Tingkat Kemiripan *Simple Matching Coefficient*

Kasus	M10	M01	M11	M00	<i>Simple Matching Coefficient</i>
Kasus 1	2	1	0	5	0.625
Kasus 2	0	4	3	1	0.5
Kasus 3	1	5	2	0	0.25

Kasus 4	0	1	6	1	0.875
Kasus 5	0	5	2	1	0.375
Kasus 6	0	5	2	1	0.375
Kasus 7	0	2	5	1	0.75
Kasus 8	0	5	2	1	0.375
Kasus 9	1	4	3	0	0.375
Kasus 10	0	5	2	1	0.375
Kasus 11	0	5	2	1	0.375
Kasus 12	1	5	2	0	0.25
Kasus 13	1	5	2	0	0.25
Kasus 14	0	4	3	1	0.5
Kasus 15	0	4	3	1	0.375
Kasus 16	0	4	3	1	0.5
Kasus 17	0	5	2	1	0.375
Kasus 18	0	4	3	1	0.5
Kasus 19	0	5	2	1	0.375
Kasus 20	1	3	4	0	0.5

Dari tabel 2 diatas terlihat bahwa kasus baru (y) lebih mirip dengan kasus 4 (kasus lama (x)) dengan tingkat kemiripan 0,875. Jadi nilai *similarity* antara kasus lama dan kasus baru adalah 0.875. Maka pada kasus baru dapat menggunakan solusi pada kasus lama yaitu pada kasus 4.

3.3 Perhitungan Tingkat Kemiripan *Sorensen-Dice Coefficient*

Dari formula *Sorensen-Dice Coefficient* dapat dirumuskan sesuai dengan representasi kasus yang terjadi pada penalaran berbasis kasus dengan x sebagai kasus lama dan y sebagai kasus baru, sehingga didapatkan formula :

$$SD(x, y) = \frac{2(M_{11})}{(2M_{11}+M_{10}+M_{01})}$$

dimana :

x : kasus lama

y : kasus baru

M_{11} : jumlah atribut biner, x=1 dan y=1

M_{10} : jumlah atribut biner, x=1 dan y=0

M_{01} : jumlah atribut biner, x=0 dan y=1

Kasus baru (y) adalah kasus yang akan dicari solusinya dengan cara membandingkan fitur gejala pada setiap kasus lama (x) atau kasus yang tersimpan di basis kasus. Kasus baru dikatakan similar (mirip) 100% dengan kasus yang lama apabila nilai similaritas dari SD(x,y) sama dengan 1 sedangkan tidak similar apabila nilai SD(x,y) sama dengan 0.

Tabel 3. Perhitungan Tingkat Kemiripan *Sorensen-Dice Coefficient*

Kasus	M10	M01	M11	M00	<i>Sorensen-Dice Coefficient</i>
Kasus 1	2	1	0	5	0
Kasus 2	0	4	3	1	0.6
Kasus 3	1	5	2	0	0.4
Kasus 4	0	1	6	1	0.923
Kasus 5	0	5	2	1	0.444
Kasus 6	0	5	2	1	0.444

Kasus 7	0	2	5	1	0.833
Kasus 8	0	5	2	1	0.444
Kasus 9	1	4	3	0	0.545
Kasus 10	0	5	2	1	0.444
Kasus 11	0	5	2	1	0.444
Kasus 12	1	5	2	0	0.4
Kasus 13	1	5	2	0	0.4
Kasus 14	0	4	3	1	0.6
Kasus 15	0	4	3	1	0.6
Kasus 16	0	4	3	1	0.6
Kasus 17	0	5	2	1	0.444
Kasus 18	0	4	3	1	0.6
Kasus 19	0	5	2	1	0.444
Kasus 20	1	3	4	0	0.667

Dari tabel 3 perhitungan tingkat kemiripan menggunakan *Sorensen-Dice Coefficient* didapatkan bahwa kasus baru memiliki tingkat kemiripan paling tinggi dengan kasus 4 (kasus lama) sebesar 0.923. Sehingga kasus baru tersebut dapat menggunakan solusi dari kasus 4.

Tabel 4. Hasil Perbandingan *Simple Matching Coefficient* dengan *Sorensen-Dice Coefficient*

Kasus	<i>Simple Matching Coefficient</i>	<i>Sorensen-Dice Coefficient</i>
Kasus 1	0.625	0
Kasus 2	0.5	0.6
Kasus 3	0.25	0.4
Kasus 4	0.875	0.923
Kasus 5	0.375	0.444
Kasus 6	0.375	0.444
Kasus 7	0.75	0.833
Kasus 8	0.375	0.444
Kasus 9	0.375	0.545
Kasus 10	0.375	0.444
Kasus 11	0.375	0.444
Kasus 12	0.25	0.4
Kasus 13	0.25	0.4
Kasus 14	0.5	0.6

Kasus 15	0.375	0.6
Kasus 16	0.5	0.6
Kasus 17	0.375	0.444
Kasus 18	0.5	0.6
Kasus 19	0.375	0.444
Kasus 20	0.5	0.667

Tabel 4 menunjukkan bahwa tingkat similaritas paling tinggi dari proses perhitungan kasus lama dengan kasus baru menggunakan *Simple Matching Coefficient* dan *Sorensen-Dice Coefficient* berada pada kasus ke 4, yaitu sebesar 0.875 dan 0.923.

Kemudian dapat dilakukan rata-rata beserta persentasenya agar menghasilkan perbandingan diantara metode similaritas tersebut.

$$\text{Rata-rata Simple Matching Coefficient} : \frac{8.875}{20} = 0.44375$$

$$\text{Rata-rata Sorensen-Dice Coefficient} : \frac{10.276}{20} = 0.5138$$

$$\text{Persentase Simple Matching Coefficient} : \frac{0.44375}{0.95755} \times 100 \% = 46.3422 \%$$

$$\text{Persentase Sorensen-Dice Coefficient} : \frac{0.5138}{0.95755} \times 100 \% = 53.6577 \%$$

Dari hasil perhitungan persentase yang dilakukan, didapatkan bahwa metode *Simple Matching Coefficient* memiliki persentase similaritas sebesar 46.3422% dan metode *Sorensen-Dice Coefficient* mempunyai persentase similaritas sebesar 53.6577%.

4. Kesimpulan dan Saran

4.1 Kesimpulan

- Dari hasil penelitian yang dilakukan diperoleh bahwa metode similaritas *Simple Matching Coefficient* dan *Sorensen-Dice Coefficient* dapat melakukan perhitungan tingkat kemiripan. Dengan hasil kemiripan antara kasus baru dengan kasus lama berada pada kasus yang sama, yaitu kasus 4 sebesar 0.875 untuk metode *Simple Matching Coefficient* dan sebesar 0.923 untuk metode *Sorensesn-Dice Coefficient*.
- Dihasilkan persentase perbandingan antara kedua metode tersebut, yaitu 46.3422% untuk metode similaritas *Simple Matching Coefficient* dan 53.6577% untuk metode similaritas *Sorensen-Dice Coefficient*.

4.2 Saran

- Dapat dilakukan pengembangan kearah pengimplementasian sistem yang dapat digunakan untuk membantu dan memecahkan kasus permasalahan yang terjadi.

- b. Dapat digunakan sample data yang lebih banyak dan lebih kompleks untuk hasil yang lebih baik.

Daftar Pustaka :

- Faizal, Edy. 2014. Case Based Reasoning Diagnosis Penyakit Cardiovascular dengan Metode Simple Matching Coefficient Similarity. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, Vol 1, No 1, pp 83-90.
- Amaliah, N., Johar, A., Dharmayana, I. W. 2017. Implementasi Metode Case-Based Reasoning (CBR) dalam Menentukan Klasifikasi anak yang Mengalami Reterdasi Mental. *Jurnal Rekursif*, Vol 5, No 1, pp 33-42.
- Aamodt, A. and Plaza, E. 1994. Case-Based Reasoning: Foundational Issues, Methodological Variations and System Approches. *AICom-Artificial Intelligence Communications. IOS Press*, Vol. 7: 1, pp.39-59.
- Tursina. 2012. Case-Based Reasoning untuk Diagnosa Penyakit Respirologi Anak Menggunakan Similaritas Simple Matching Coefficient. *Jurnal ELKHA*, Vol 4, No 1, Halaman 17-22.
- Chao, A., Chazdon, R. L., Colwell, R. K., and Shen, T. 2006. Abundance-Based Similarity Indices and Their Estimation When There Are Unseen Species in Samples. *BIOMETRICS* 62, Pp 361-371.