

PERBANDINGAN *JACCARD SIMILARITY* DENGAN *EXTENDED JACCARD SIMILARITY* PADA PENALARAN BERBASIS KASUS

Murien Nugraheni¹

¹ Prodi Sistem dan Teknologi Informasi, FT – UNJ
¹muriennugraheni@unj.ac.id

Abstrak

Penalaran berbasis kasus memberikan solusi pada suatu kasus baru dengan menggunakan pengetahuan lalu yang tersimpan pada basis pengetahuan. Dilakukan perhitungan kasus baru dengan menggunakan sampel gejala yang akan didiagnosa dengan kasus lama yang telah tersimpan pada basis data. Kasus lama dengan nilai similaritas tertinggi akan diambil dan kemudian solusi dari kasus lama tersebut akan dijadikan solusi bagi kasus yang baru. Jika suatu kasus tidak berhasil didiagnosa, maka kasus tersebut akan dilakukan revisi oleh seorang pakar. Kasus yang berhasil direvisi akan disimpan ke dalam basis data untuk dijadikan pengetahuan baru. Pada penelitian ini dilakukan dua perbandingan metode untuk menghitung tingkat kemiripan antara kasus baru dengan kasus lama. Metode similarity yang digunakan adalah metode Jaccard Similarity dan Extended Jaccard Similarity. Hasil dari penelitian ini diperoleh bahwa metode Extended Jaccard Similarity mempunyai tingkat kemiripan sebesar 52.6818% sedangkan metode Jaccard Similarity memiliki tingkat kemiripan sebesar 47.3181%.

Kata kunci : Case-Based Reasoning, Jaccard Similarity, Extended Jaccard Similarity

1. Pendahuluan

Penalaran berbasis kasus merupakan penalaran komputer yang menggunakan pengetahuan lama untuk mengatasi permasalahan baru. Penalaran berbasis kasus memberikan solusi terhadap kasus baru dengan melihat kasus lama yang mempunyai kemiripan paling mendekati dengan kasus baru. Hal ini akan sangat bermanfaat karena dapat menghilangkan kebutuhan untuk mengekstrak model seperti yang dibutuhkan oleh sistem berbasis aturan. Selain itu, penalaran berbasis kasus juga dapat dimulai dari jumlah pengetahuan yang sedikit karena pengetahuan penalaran berbasis kasus dapat bertambah secara bertahap ketika sebuah kasus ditambahkan.

Proses mencari kedekatan atau kemiripan antara kasus baru dengan kasus lama untuk memperoleh solusi terhadap kasus baru pada penalaran berbasis kasus dapat menggunakan berbagai macam metode, dimana metode ini akan mempengaruhi keberhasilan dari penalaran berbasis kasus. Beberapa metode yang dapat digunakan dalam mencari kedekatan atau kemiripan kasus baru dengan kasus lama adalah *Jaccard Similarity* dan *Extended Jaccard Similarity*. Metode ini mengkorelasikan dua objek variabel kontinue yang mempunyai hubungan linier diantara atribut pada suatu objek.

Penelitian yang dilakukan oleh Fatmayati (2017:70) membandingkan dua metode *Extended Jaccard Coefficient (Tanimoto Coefficient)* dan *Euclidean Distance Similarity* untuk mendiagnosa penyakit gigi dan mulut. Dengan menghasilkan

tingkat akurasi sebesar 95,24% untuk *Extended Jaccard Coefficient* dan tingkat akurasi sebesar 100% untuk *Euclidean Distance Similarity*.

Nurdiana (2016:59) dalam penelitiannya membandingkan metode *Cosine Similarity* dengan *Jaccard Similarity* pada aplikasi pencarian terjemah Al-Qur'an dalam Bahasa Indonesia. Dari hasil penelitian yang dilakukan, metode *Cosine Similarity* menghasilkan kemiripan tertinggi sebesar 41% dari metode *Jaccard Similarity* yang hanya 19%.

2. Metode

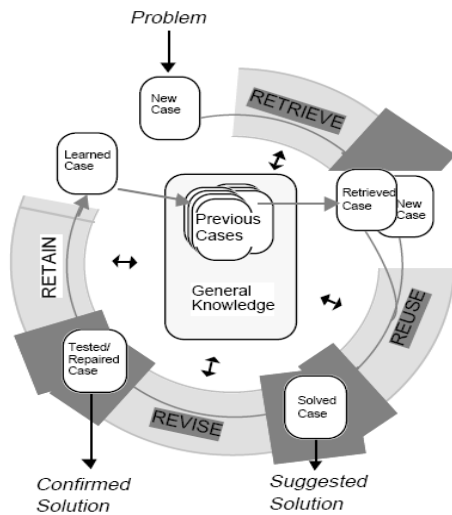
2.1 Penalaran Berbasis Kasus

Penalaran berbasis kasus merupakan metode dari sistem pendukung keputusan yang menghasilkan solusi yang dibutuhkan dengan persamaan dari pengalaman yang terdahulu. Metode ini terdiri dari mengacu kembali, menggunakan kembali, meninjau ulang dan mendalami kasus yang telah lalu. Memproses permasalahan yang diajukan dengan menggunakan solusi pada kasus sebelumnya yang memiliki persamaan. Proses tersebut akan menghasilkan solusi yang telah dikembangkan dan disesuaikan untuk mengatasi permasalahan.

Penalaran berbasis kasus dapat direpresentasikan sebagai suatu siklus proses yang dibagi menjadi empat sub proses (Aamodt dan Plaza, 1994), yaitu :

a. *Retrieve* yaitu mencari kasus-kasus sebelumnya yang paling mirip dengan kasus baru.

- b. *Reuse* yaitu menggunakan kembali kasus-kasus yang paling mirip tersebut untuk mendapatkan solusi untuk kasus yang baru.
- c. *Revise* yaitu melakukan penyesuaian dari solusi kasus-kasus sebelumnya agar dapat dijadikan solusi untuk kasus yang baru.
- d. *Retain* yaitu memakai solusi baru sebagai bagian dari kasus baru, kemudian kasus baru di-update ke dalam basis kasus.



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

Penalaran berbasis kasus menitikberatkan pemecahan masalah dengan didasarkan pada *knowledge* dari kasus-kasus sebelumnya, apabila ada kasus baru maka akan disimpan pada basis pengetahuan sehingga sistem akan melakukan *learning* dan *knowledge* yang dimiliki oleh sistem akan bertambah. Secara umum metode ini terdiri dari 4 langkah yaitu :

1) *Retrieve*

Pada proses ini adalah proses pengenalan masalah, yaitu dengan mengidentifikasi masalah yang baru.

2) *Reuse*

Dalam proses *Reuse*, sistem akan melakukan pencarian masalah terdahulu pada *database* melalui identifikasi masalah baru. Kemudian menggunakan kembali informasi permasalahan terdahulu tersebut yang memiliki kesamaan untuk menyelesaikan permasalahan yang baru.

3) *Revise*

Pada proses ini, informasi akan dievaluasi kembali untuk mengatasi masalah yang terjadi pada permasalahan baru. Kemudian sistem akan mengeluarkan solusi masalah baru.

4) *Retain*

Pada proses terakhir ini, sistem akan menyimpan permasalahan baru ke dalam *knowledge-base* untuk menyelesaikan permasalahan yang akan datang.

2.2 Jaccard Similarity

Salah satu cara untuk menghitung similaritas dua objek (items) yang bersifat biner adalah dengan menggunakan metode *Jaccard Similarity*.

Formula yang digunakan *Jaccard Similarity* untuk menghitung similaritas antara dua objek x dan y adalah sebagai berikut :

$$J(x,y) = \frac{M_{11}}{M_{01} + M_{10} + M_{11}} \dots \dots \text{(Tan, Steinbach, Kumar,2006)}$$

Keterangan :

- x : kasus lama
- y : kasus baru
- M_{11} : jumlah atribut biner dimana x=1 dan y=1
- M_{10} : jumlah atribut biner dimana x=1 dan y=0
- M_{01} : jumlah atribut biner dimana 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 *Jaccard(x,y)* sama dengan 1 sedangkan tidak similar apabila nilai *Jaccard(x,y)* sama dengan 0.

2.3 Extended Jaccard Similarity

Extended Jaccard Coefficient atau yang sering disebut sebagai *Tanimoto Coefficient* adalah salah satu metode similaritas yang digunakan untuk melakukan perhitungan tingkat kemiripan dari dua buah objek.

Extended Jaccard Coefficient dapat di rumuskan sebagai berikut :

$$EJ(x,y) = \frac{x \cdot y}{||x||^2 + ||y||^2 - x \cdot y} \dots \dots \text{(Tan, Steinbach, Kumar, 2006)}$$

Dimana x adalah kasus lama dan y adalah kasus baru dan “|| ||” menunjukkan *norm* pada masing-masing vektor.

3. Hasil dan Pembahasan

3.1 Representasi Kasus

Secara umum, suatu kasus direpresentasikan dalam bentuk kumpulan fitur-fitur yang menjadi ciri kasus tersebut dan solusi untuk menangani kasus tersebut. Fitur-fitur yang disimpan merupakan parameter untuk mendapatkan solusi. Fitur-fitur tersebut bisa didapat dari akuisisi pengetahuan, dengan metode pengumpulan data studi pustaka.

Sebagai contoh pada perbandingan tingkat kemiripan suatu kasus dengan menggunakan dua metode *similarity* yang berbeda yaitu metode *Jaccard Similarity* dan metode *Extended Jaccard Similarity* dengan menggunakan 20 kasus yang telah tersimpan di dalam basis data untuk dilakukan perhitungan tingkat kemiripannya.

Proses dimulai dengan melakukan perhitungan *similarity* kasus baru terhadap kasus lama. Proses perhitungan *similarity* menggunakan metode *Jaccard Similarity* dan *Extended Jaccard Similarity*. Nilai *similarity* yang digunakan yaitu 0 dan 1. Dimana 0 menunjukkan tidak terlihat gejala yang muncul sedangkan nilai 1 menunjukkan bahwa gejala terlihat.

3.2 Perhitungan Similaritas

Terdapat 20 kasus dengan 9 gejala pada kasus lama yang tersimpan di dalam *database*.

Tabel 1. Contoh Representasi Kasus

K	G1	G2	G3	G4	G5	G6	G7	G8	G9
1	1	1	0	0	0	1	0	0	0
2	1	0	1	0	0	1	0	0	0
3	1	0	0	0	1	1	0	0	0
4	1	0	0	0	0	0	1	0	0
5	1	1	0	0	0	0	1	0	0
6	1	0	1	0	0	0	1	0	0
7	1	0	0	1	0	0	1	0	0
8	1	0	0	0	1	0	1	0	0
9	0	0	0	0	0	1	0	1	0
10	0	0	0	0	0	1	0	0	1
11	1	0	0	0	0	1	0	0	0
12	0	1	0	0	0	1	0	0	0
13	0	0	1	0	0	1	0	0	0
14	0	0	0	1	0	1	0	0	0
15	0	0	0	0	1	1	0	0	0
16	0	1	0	0	0	0	1	0	0
17	0	0	1	0	0	0	1	0	0
18	0	0	0	1	0	0	1	0	0
19	0	0	0	0	1	0	1	0	0
20	0	0	0	0	0	0	1	0	0

Tabel 2. Contoh Kasus Baru

	G1	G2	G3	G4	G5	G6	G7	G8	G9
y	1	0	0	1	0	1	0	0	0

Akan diselesaikan menggunakan rumus untuk *Jaccard Similarity* sebagai berikut :

$$Jaccard (M) = \frac{M_{11}}{M_{01} + M_{10} + M_{11}}$$

dimana :

- x : kasus lama
- y : kasus baru
- M_{11} : jumlah atribut biner dimana $x=1$ dan $y=1$
- M_{10} : jumlah atribut biner dimana $x=1$ dan $y=0$
- M_{01} : jumlah atribut biner dimana $x=0$ dan $y=1$

Sedangkan rumus *Extended Jaccard Similarity* adalah sebagai berikut :

$$EJ(x,y) = \frac{x \cdot y}{||x||^2 + ||y||^2 - x \cdot y}$$

Contoh perhitungan kasus 1 :

Kasus 1 merupakan nilai dari kasus lama (x) yang dibandingkan dengan kasus baru (y).

Tabel 3. Contoh Kasus 1

Kasus 1	1	1	0	0	0	1	0	0	0
Kasus Baru (y)	1	0	0	1	0	1	0	0	0

Maka diperoleh :

$$M_{01} = 1; M_{10} = 1; M_{11} = 2$$

Sehingga

$$Jaccard (x,y) = 2 / (1 + 1 + 2) = 2 / 4 = 0.5$$

Untuk *Extended Jaccard*

$$\begin{aligned} (x,y) &= [(x_1*y_1) + [(x_2*y_2) + [(x_3*y_3) + [(x_4*y_4) + \\ & [(x_5*y_5) + [(x_6*y_6) + [(x_7*y_7) + [(x_8*y_8) + \\ & [(x_9*y_9)]]]]]]]] \\ &= [(1*1) + (1*0) + (0*0) + (0*1) + (0*0) + \\ & (1*1) + (0*0) + (0*0) + (0*0)] \\ &= 1 + 0 + 0 + 0 + 0 + 1 + 0 + 0 + 0 \\ &= 2 \end{aligned}$$

$$\begin{aligned} ||x|| &= \sqrt{x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 + x_6^2 + x_7^2 + x_8^2 + x_9^2} \\ &= \sqrt{1^2 + 1^2 + 0^2 + 0^2 + 0^2 + 1^2 + 0^2 + 0^2 + 0^2} \\ &= \sqrt{2} \end{aligned}$$

$$||x||^2 = (\sqrt{2})^2 = 2$$

$$\begin{aligned} ||y|| &= \sqrt{y_1^2 + y_2^2 + y_3^2 + y_4^2 + y_5^2 + y_6^2 + y_7^2 + y_8^2 + y_9^2} \\ &= \sqrt{1^2 + 0^2 + 0^2 + 1^2 + 0^2 + 1^2 + 0^2 + 0^2 + 0^2} \\ &= \sqrt{3} \end{aligned}$$

$$||y||^2 = (\sqrt{3})^2 = 3$$

$$EJ(x,y) = 2 / 2 + 3 - 2 = 2 / 3 = 0.667$$

Hasil perhitungan similaritas dengan menggunakan *Jaccard Similarity* dan *Extended Jaccard Similarity* adalah sebagai berikut :

Tabel 4. Hasil Perhitungan Similaritas

	<i>Jaccard Similarity</i>	<i>Extended Jaccard Similarity</i>
Kasus 1	0.5	0.667
Kasus 2	0.5	0.667
Kasus 3	0.5	0.667
Kasus 4	0.25	0.25
Kasus 5	0.25	0.25
Kasus 6	0.25	0.25
Kasus 7	0.5	0.667
Kasus 8	0.25	0.25
Kasus 9	0.25	0.25
Kasus 10	0.25	0.25
Kasus 11	0.67	0.667
Kasus 12	0.25	0.25
Kasus 13	0.25	0.25
Kasus 14	0.67	0.667
Kasus 15	0.25	0.25
Kasus 16	0	0
Kasus 17	0	0
Kasus 18	0.25	0
Kasus 19	0	0.25
Kasus 20	0	0

Dari perhitungan 20 kasus didapatkan rata-ratanya yang kemudian dapat dilakukan perhitungan persentasenya untuk mendapatkan hasil perbandingan diantara kedua metode.

Rata-rata *Jaccard Similarity* : $\frac{5.84}{20} = 0.292$ dan rata-rata *Extended Jaccard Similarity* : $\frac{6.502}{20} = 0.3251$

Persentase *Jaccard Similarity* :

$$\frac{0.292}{0.6171} \times 100 \% = 47.3181 \%$$

Persentase *Extended Jaccard Similarity* :

$$\frac{0.3251}{0.6171} \times 100 \% = 52.6818 \%$$

Metode *Jaccard Similarity* mempunyai persentase tingkat kemiripan sebesar 47.3181% dan metode *Extended Jaccard Similarity* mempunyai persentase tingkat kemiripan sebesar 52.6818%.

5. Kesimpulan dan Saran

5.1 Kesimpulan

Dari hasil penelitian yang dilakukan didapat bahwa metode *Extended Jaccard Similarity* menghasilkan tingkat similaritas lebih tinggi dibandingkan dengan metode *Jaccard Similarity* yaitu sebesar 52.6818 % untuk metode *Extended Jaccard Similarity* dan sebesar 47.3181 % untuk metode *Jaccard Similarity*.

5.3 Saran

Dari hasil penelitian ini diharapkan dapat dilakukan :

- 1 Perlu digunakan sampel data yang lebih banyak dan kompleks agar menghasilkan tingkat similaritas yang lebih akurat.
- 2 Perlu dilakukan perbandingan dengan menggunakan metode lain untuk menghasilkan tingkat kemiripan yang lebih tinggi.

Daftar Pustaka:

- Fatmayati, Fryda. (2017). *Implementasi Case Base Reasoning untuk Mendiagnosa Penyakit Gigi dan Mulut*. Techno.COM. Vol. 16 No. 1. pp. 70-79. Semarang.
- Nurdiana, O., Jumadi, dan Nursantika, D. (2016). *Perbandingan Metode Cosine Similarity dengan Metode Jaccard Similarity pada Aplikasi Pencarian Terjemah Al-Qur'an dalam Bahasa Indonesia*. Jurnal Online Informatika. Vol. 1 No. 1. pp. 59-63. Bandung.
- 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.

Tan, P., Steinbach, M., Kumar, V. (2006). *Intoduction To Data Mining*. Boston: Pearson Education Inc.

Available at:

<http://journal.unj.ac.id/unj/index.php/pinter/article/view/18968>