

# PENERAPAN ROUGH SET QUANTITATIVE MEASURE PADA APLIKASI PENDUKUNG KEPUTUSAN

**Rully Soelaiman<sup>1)</sup>, Wiwik Anggraeni, Eko Setiawan**

Fakultas Teknologi Informasi, Institut Teknologi Sepuluh Nopember,  
Kampus Keputih, Sukolilo, Surabaya 60111, Indonesia  
Email : rully@is.its.ac.id<sup>1)</sup>

## Abstraksi

*Teori Rough set membagi data ke dalam kelas upper approximation, lower approximation, dan boundary. Daerah rough (boundary) dalam klasifikasi Rough set dapat didekati dengan perhitungan quantitative measure. Kombinasi cara tersebut dengan memakai query SQL merupakan hal yang baru.*

*Teori Rough set akan diimplementasikan dan dikembangkan dengan dasar query SQL. Dengan Rough set, dimungkinkan untuk menemukan hubungan tersembunyi dan reduct atribut dari serangkaian klasifikasi data atributif. Dari reduct tersebut dapat digenerate rules, dan dihitung quantitative measures untuk memperkuat rule tersebut. Rule yang diperkuat quantitative measure digunakan untuk mengklasifikasikan data.*

*Dalam pengembangan ini dihasilkan aplikasi medis diagnosa penyakit Medirose. Selanjutnya diberikan contoh penggunaan Medirose dengan beberapa contoh kasus untuk mengilustrasikan kegunaan dari sistem Rough set dan quantitative measure ini.*

*Kata kunci : Rough set, quantitative measure.*

## PENDAHULUAN

Teori *Rough set* telah dikembangkan oleh Zdzislaw Pawlak pada awal tahun 1980-an. Ini merupakan pengembangan dari teori *fuzzy set* (Lotfi A. Zadeh, 1965) yang mengatakan adanya daerah abu-abu (antara hitam dan putih) dimana suatu hal tidak hanya direpresentasikan sebagai *binary*, *boolean*, ataupun benar-salah saja. Dalam teori matematika, keanggotaan dalam suatu himpunan didefinisikan untuk setiap obyek dalam semesta yang dapat dikenali, dan sisanya merupakan milik komplemen dari himpunan yang dimaksud. Pada kenyataannya, informasi obyek yang tersedia sering tidak cukup untuk mendefinisikan obyek dan melakukan klasifikasi. Teori rough set membuat analisis formal agar situasi yang demikian dapat diatasi. Tujuan dari teori *Rough set* adalah untuk mengenali ketidakpastian dalam klasifikasi dari suatu obyek. Vitoria, Dam'asio dan Maluszy'nski (2004) mengembangkan bahasa untuk meng-ekspresikan data *Rough* dan dilengkapi dengan fitur *quantitative measure* seperti *support*, *strength* dan *accuracy*.

### *Rough Set Quantitative Measure*

Dalam analisa data, suatu obyek dapat direpresentasikan sebagai sistem informasi. **Sistem informasi** adalah pasangan  $I = (U, A)$ , dimana  $U$  adalah bukan himpunan kosong terbatas dari obyek,

biasanya disebut semesta, dan  $A$  adalah bukan himpunan kosong terbatas dari himpunan atribut yang dimiliki elemen-elemen dari  $U$ . Informasi mengenai obyek (elemen) direpresentasikan dengan sebuah himpunan atribut yang memiliki nilai. Sebuah atribut  $\alpha \in A$  adalah bagian dari fungsi  $\alpha : U \rightarrow V_\alpha$ ,

dimana  $V_\alpha$  adalah himpunan nilai untuk  $\alpha$ .

Pawlak (2004) menyatakan bahwa sistem keputusan adalah sebuah sistem informasi yang diperluas dengan sebuah himpunan atribut keputusan (*decision attribute*). **Sistem keputusan** adalah sebuah sistem informasi  $I$  yang disampaikan dengan sebuah himpunan atribut keputusan  $D$ , sehingga  $I = (U, A, D)$  dimana  $D \cap A = \emptyset$ . Dalam sebuah sistem keputusan, atribut  $A$  disebut atribut kondisional (*conditional attribute*).

Tabel 1: Tabel keputusan

	headache	muscle pain	temperature	flu
p1	no	yes	high	yes
p2	yes	no	high	yes
p3	yes	yes	very high	yes
p4	no	yes	normal	no
p5	yes	no	high	no
p6	no	yes	very high	yes

Misalkan  $I = (U, A, D)$  adalah sebuah sistem keputusan. Setiap  $d_i \in D$  membagi semesta  $U$  ke

dalam  $|V_{d_i}|$  kelas  $X_1, \dots, X_k$ . Setiap kelas  $X_j (j \in \{1, \dots, |V_{d_i}|\})$  disebut **decision class**.

Obyek yang memiliki nilai yang sama untuk atribut kondisionalnya disebut *indiscernible*. Contohnya: pasien yang memiliki sekumpulan gejala yang sama seperti pasien p2 dan p5 pada tabel 1. Menurut Komorowski dan Øhrn (1999), teori *Rough set* menempatkan kedua pasien tersebut ke dalam *indiscernibility relation* yang digunakan untuk menjelaskan fakta bahwa barangkali tidak mungkin untuk memisahkan obyek tertentu dalam semesta menggunakan informasi (atribut-atribut) yang ada.

Jika  $B \subseteq A$ , *indiscernibility relation*  $IND_I(B)$  didefinisikan sebagai:

$$IND_I(B) = \{(x, x') \in U^2 \mid \forall \alpha \in B, \alpha(x) = \alpha(x')\}$$

Jika  $(x, x') \in IND_I(B)$ , maka  $x$  dan  $x'$  *indiscernible* dengan mematuhi semua atribut dalam  $B$ . Hubungan seperti diatas disebut *equivalence relations*. *Equivalence class* dari sebuah obyek  $x \in U$  terdiri dari semua obyek  $y \in U$  sehingga  $(x, y) \in IND(B)$ . *Equivalence class* diperoleh dari  $IND(B)$  yang dinotasikan  $[x]_B$ , dengan  $x \in U$ .

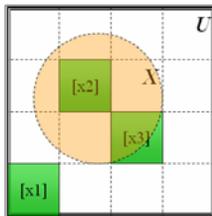
Dari informasi di tabel 1 dapat dilihat bahwa:

$$IND(\{headache\}) = \{\{p1, p4, p6\}, \{p2, p3, p5\}\},$$

$$IND(\{musclepain\}) = \{\{p1, p3, p4, p6\}, \{p2, p5\}\},$$

...

$$IND(\{headache, musclepain, temperature\}) = \{\{p1\}, \{p2, p5\}, \{p3\}, \{p4\}, \{p6\}\}.$$



Gambar 1 : Equivalence class dalam semesta U dan hubungannya dengan concept X.

Berdasarkan gambar 1,  $U$  adalah semesta,  $X \subset U$  adalah sebuah himpunan bagian yang disebut *concept*, dan  $[x_i]_A$  *equivalence class* yang dibentuk dengan  $IND_I(A)$ . Setiap *equivalence class* pada gambar 1 diwakili oleh persegi. *Equivalence class*  $[x2]$  berada dalam *concept*  $X$ , yang artinya obyek dari  $[x2]$  merupakan anggota dari  $X$ . Sedangkan  $[x1]$  diluar *concept*  $X$ , berarti bukan anggota *concept*  $X$ . Kasus yang problematik datang dari ambiguitas *equivalence class*  $[x3]$  yang sebagian didalam dan sebagian lagi diluar *concept*  $X$ . *Rough set* dapat menyelesaikan kasus seperti ini dengan pendekatan himpunan klasik untuk

menunjukkan anggota lain yang merupakan anggota yang mungkin atau pasti.

Misalkan  $I = (U, A, D)$  merupakan sistem keputusan,  $B \subseteq A$ , dan  $X \subseteq U$ .

- *Lower Approximation* himpunan  $\underline{B}(X)$  dan didefinisikan sebagai:

$$\underline{B}(X) = \{x \in U \mid [x]_B \subseteq X\}.$$

- *Upper Approximation* himpunan  $\overline{B}(X)$  didefinisikan sebagai:

$$\overline{B}(X) = \{x \in U \mid [x]_B \cap X \neq \emptyset\},$$

- *Boundary* didefinisikan dengan:

$$\overline{B}(X) = \overline{B}(X) - \underline{B}(X).$$

*Boundary* merupakan daerah perbatasan, yaitu selisih dari *Upper Approximation* dan *Lower Approximation*.

Jika  $\overline{B}(X) = \emptyset$  maka  $X$  adalah pasti (*crisp*) dan jika

$\overline{B}(X) \neq \emptyset$  maka  $X$  adalah *rough* (tidak pasti). Daerah *boundary*  $\overline{B}(X)$  menunjukkan ambiguitas di dalam informasi mengenai obyek di dalam  $X$  dan semua obyek yang tidak konsisten dalam *concept*  $X$ . Untuk selanjutnya kita akan notasikan  $\underline{B}(X)$  dengan  $\underline{X}$ ,  $\overline{B}(X)$  dengan  $\overline{X}$ , dan  $\overline{B}(X)$  dengan  $\overline{\overline{X}}$ .

Sebuah *Rough set*  $S$  didefinisikan sebagai pasangan

$$S = (\underline{A}(S), \overline{A}(S)), \text{ dimana } \neg S = U - S.$$

Berikut ini merupakan hasil pendekatan *Rough set* pada data pasien flu:

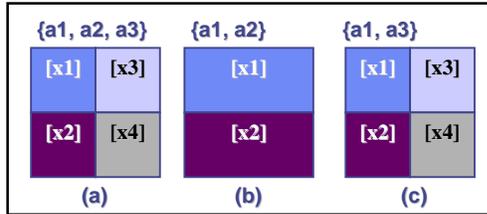
$$\underline{X} = \{p1, p3, p6\},$$

$$\overline{X} = \{p1, p2, p3, p5, p6\},$$

$$\overline{\overline{X}} = \{p4\},$$

$$\overline{\overline{\overline{X}}} = \{p2, p4, p5\}.$$

Misalkan  $I = (U, A)$  merupakan sebuah sistem informasi dan  $A = \{a1, a2, a3\}$ . Gambar 2 akan mengilustrasikan bagai-mana pemilihan dari atribut di dalam  $A$  merubah pembagian  $U$  kedalam *equivalence class* yang berbeda. Pada gambar 2 dapat dilihat bahwa himpunan atribut  $\{a1, a3\}$  dan  $A$  menghasilkan beberapa beberapa himpunan *equivalence class* yang sama. Ini berarti bahwa atribut  $a2$  tidak dibutuhkan untuk membedakan obyek dalam  $U$ . Dengan demikian ada kebutuhan untuk mendapatkan jumlah minimal dari atribut yang tetap menjaga *indiscernibility relation*.



Gambar 2: Indiscernibility Relation: (a) Semua atribut original di dalam A digunakan, menghasilkan 4 *equivalence class* yang berbeda. (b) Atribut a3 dihapus dari atribut set A, menghasilkan hanya 2 *equivalence class*. (c) Atribut a2 dihapus menghasilkan 4 *equivalence class* yang berbeda (sama seperti (a)).

Menurut Pawlak (2004), pengurangan atribut seperti ini disebut *reduct*, yaitu menjaga pembagian dari semesta dengan menggunakan atribut yang lebih sedikit dan tidak merubah klasifikasi obyek ketika dibandingkan dengan himpunan atribut aslinya. **Reduct** adalah himpunan minimal dari atribut  $B \subseteq A$  sehingga

$$IND_I(B) = IND_I(A).$$

Sayangnya cara untuk menemukan *reduct*, seperti yang telah diteliti secara seksama oleh Andersson (2004), adalah sangat sulit. Meskipun demikian, ada beberapa algoritma yang relatif cepat dimana algoritma tersebut adalah bersifat algoritma *heuristic*.

Sebuah baris pada tabel keputusan dapat dilihat sebagai *decision rule*. *Decision rule* adalah sebuah pernyataan *if then*, berbentuk *if f (antecedent) then g (conclusion)*, direpresentasikan sebagai  $f \rightarrow g$ .

Beberapa *quantitative measure* biasanya dihubungkan dengan *decision rule*, meliputi *support*, *strength*, *accuracy*, dan *coverage* yang akan dijelaskan berikut ini.

*Support* dari *decision rule* adalah jumlah obyek yang baik *antecedent* maupun *conclusion*-nya sesuai. *Support* dari *decision rule*  $f \rightarrow g$  di dalam  $I$  didefinisikan sebagai

$Support(f \rightarrow g) = card(\| f \|_I \cap \| g \|_I)$  dimana  $card$  menotasikan jumlah anggota himpunan (kardinalitas).

*Strength* dari *decision rule* mengindikasikan berapa sering obyek dalam semesta memenuhi rule.

$$Strength(f \rightarrow g) = \frac{Support(f \rightarrow g)}{card(U)}$$

*Accuracy* menunjukkan rasio perbandingan obyek yang memenuhi *antecedent* dan juga memenuhi *conclusion* terhadap obyek yang memenuhi *antecedent* saja

$$Accuracy(f \rightarrow g) = \frac{Support(f \rightarrow g)}{card(\| f \|_I)}$$

*Coverage* merupakan rasio perbandingan obyek yang memenuhi *rule conclusion* dan juga memenuhi

*antecedent* terhadap obyek yang memenuhi *conclusion* saja.

$$Coverage(f \rightarrow g) = \frac{Support(f \rightarrow g)}{card(\| g \|_I)}$$

## UJI COBA

Pada pengembangan ini dihasilkan aplikasi Medirose yang digunakan untuk mendiagnosa pasien dengan data atributif menggunakan konsep *Rough set quantitative measure*. Data uji adalah data pasien flu yang merupakan hasil perluasan dari contoh kasus data pasien flu sebelumnya yang diperluas lagi dengan satu atribut kondisional yaitu atribut pilek, sehingga atribut kondisionalnya menjadi {Pusing, Nyeri\_otot, Temperatur, Pilek} dan jumlah data diperbanyak sampai 66 pasien. 60 data digunakan untuk training dan sisanya untuk testing sehingga akan lebih mudah untuk menunjukkan kegunaan fitur *quantitative measure*.

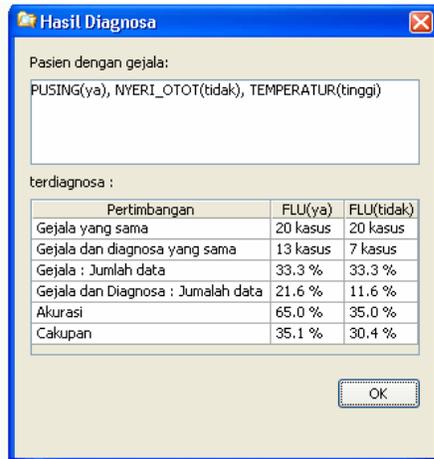
Dari uji coba yang dilakukan didapatkan *reduct* seperti pada gambar 3. *Reduct* tersebut digunakan untuk meng-generate rule yang digunakan untuk klasifikasi (diagnosa) pasien seperti contoh pada gambar 4. Testing dilakukan dengan mendiagnosa enam data pasien testing dengan menggunakan beberapa gejalanya dan membandingkan hasilnya dengan atribut keputusannya.

Pemilihan kombinasi gejala untuk tes diagnosa pada tabel 2 adalah gejala yang tercetak tebal. Pemilihan kombinasi gejala pada diagnosa ini diambil secara acak. Ini dilakukan untuk menunjukkan fungsi dari *reduct*. Sehingga untuk mendiagnosa, tidak perlu digunakan seluruh gejala yang ada. Cukup dengan kombinasi gejala yang ada pada *reduct*.

Dari keenam data uji didapatkan hasil seperti pada tabel 2, ternyata ada dua pasien yang didiagnosa antara flu dan tidak. Ini menunjukkan konsep *Rough set* yang dapat didekati dengan *quantitative measure*.

No	Gejala	Jumlah
1	TEMPERATUR, PILEK	2
2	NYERI_OTOT, PILEK	2
3	NYERI_OTOT, TEMPERATUR	2
4	NYERI_OTOT, TEMPERATUR, PILEK	3
5	PUSING, PILEK	2
6	PUSING, TEMPERATUR	2
7	PUSING, TEMPERATUR, PILEK	3
8	PUSING, NYERI_OTOT, PILEK	3
9	PUSING, NYERI_OTOT, TEMPERATUR	3
10	PUSING, NYERI_OTOT, TEMPERATUR, PILEK	4

Gambar 3: Kombinasi gejala minimal (*reduct*) untuk diagnosa



Gambar 4: Contoh diagnosa pasien flu

### SIMPULAN

Dari beberapa percobaan yang telah dilakukan dalam implementasi program untuk menerapkan teori Rough set didapatkan simpulan sebagai berikut :

1. Teori *Rough set* dapat diterapkan dengan menggunakan queri SQL pada database dengan aturan dan cara yang telah dijelaskan.
2. Klasifikasi *equivalent class* yang *rough* atau *inconsistent* dapat didekati dengan *quantitative measure*.

3. Diagnosis penyakit pada pasien dengan gejala atributif dapat dilakukan dengan pendekatan *Rough set*.
4. Aplikasi *Medirose* yang dihasilkan dari penelitian ini dapat membantu medis dalam mendiagnosis penyakit yang memiliki gejala-gejala atributif.

### DAFTAR PUSTAKA

- [1] Andersson, R. (2004). Implementation of a Rough Knowledge Base System Supporting Quantitative Measures. Linkoping: Linkoping University Electronic Press.
- [2] Komorowski, J., Øhrn, A. (1999). Modelling Prognostic Power of Cardiac Tests Using Rough Sets. *Artificial Intelligence in Medicine*, 15(2).167-191. Elsevier.
- [3] Pawlak, Z. (2004). Some Issues on Rough Sets. J.F. Peters et al. (Eds.): *Transactions on Rough Sets I*, LNCS 3100, 1–24, Berlin Heidelberg: Springer-Verlag.
- [4] Vitoria, A., Dam´asio, C.V., & Maluszy´nski, J. (2003). From Rough sets to Rough knowledge bases. *Fundamenta Informaticae*, 1-32. IOS Press.
- [5] Vitoria, A., Dam´asio, C.V., & Maluszy´nski, J. (2004). Toward Rough knowledge bases with quantitative measures. *Rough Sets and Current Trends in Computing Volume 3066 of Lecture Notes in Computer Science*, 153-158. Springer.

Tabel 2: Hasil klasifikasi data uji pasien flu.

No	Pusing	Nyeri otot	Temp.	Pilek	Flu	Hasil (Flu)	L Supp	R Supp	L Streng	R Streng	Acc.	Cov.
1	tidak	ya	tinggi	ya	ya	ya	17	17	28.3 %	28.3 %	100 %	45.9 %
2	tidak	ya	normal	tidak	tidak	tidak	16	16	26.6 %	26.6 %	100 %	69.5 %
3	ya	tidak	tinggi	tidak	ya	ya	20	13	33.3 %	21.6 %	65 %	35.1 %
						tidak				11.6 %		
4	tidak	ya	sangat tinggi	ya	ya	ya	11	11	18.3 %	18.3 %	100 %	29.7 %
5	ya	ya	sangat tinggi	ya	ya	ya	11	11	18.3 %	18.3 %	100 %	29.7 %
6	ya	tidak	tinggi	tidak	tidak	ya	20	13	33.3 %	21.6 %	65 %	35.1 %
						tidak				11.6 %		