

Seleksi *Rule* Menggunakan *Rough Set Theory* Untuk Diagnosis Penyakit Tuberkulosis

Suhardi¹, Noor Akhmad Setiawan², Indriana Hidayah³

Jurusan Teknik Elektro dan Teknologi Informasi Universitas Gadjah Mada Yogyakarta¹
Suhardi_s2te12@mail.ugm.ac.id

Jurusan Teknik Elektro dan Teknologi Informasi Universitas Gadjah Mada Yogyakarta²
Jurusan Teknik Elektro dan Teknologi Informasi Universitas Gadjah Mada Yogyakarta³

Abstrak

Tujuan dari penelitian ini adalah mengembangkan metode seleksi *rule* untuk memfilter sejumlah *rule* yang berjumlah besar untuk diagnosis penyakit Tuberkulosis menggunakan *Rough Sets Theory (RST)* pada data set Tuberkulosis. Seleksi *rule* pada penelitian ini terdiri dari dua tahapan. Ekstraksi *rule* berdasarkan batasan nilai *support* tertentu diterapkan pada tahap pertama. Seleksi berbasis *rough set* dengan konsep reduksi atribut digunakan pada tahap kedua dengan *rule* ekstrakstraksi pada tahap pertama disusun membentuk tabel keputusan baru. Dari percobaan yang telah dilakukan terhadap *rule* yang terseleksi menunjukkan bahwa metode yang diusulkan dapat memilih sejumlah kecil *rule* dengan tetap menjaga kualitas nilai klasifikasi dan mempermudah dalam pengambilan keputusan.

Kata Kunci: Ekstraksi, klasifikasi, reduksi atribut, *support*.

1. Pendahuluan

Paru-paru adalah organ penting yang merupakan salah satu organ vital bagi kehidupan manusia. Organ ini memiliki peranan pada sistem pernapasan, karena dapat memenuhi kebutuhan tubuh akan oksigen. Organ yang terletak di bawah tulang rusuk ini mempunyai tugas yang berat untuk mengatur pertukaran oksigen. Jika udara yang kita hirup tercemar dan terdapat berbagai bibit penyakit yang berkeliaran di udara akan menimbulkan berbagai penyakit paru-paru. *Tuberculosis (TB)* atau biasa disebut TBC adalah contoh penyakit yang banyak diderita oleh manusia. TB adalah penyakit menular langsung yang disebabkan oleh kuman yang disebut *Mycobacterium tuberculosis*. Sebagian besar TB menyerang paru, tetapi dapat juga mengenai organ tubuh lainnya [1]. Diperkirakan angka kematian akibat TB adalah 8000 setiap hari dan 2 - 3 juta setiap tahun [2]. Berdasarkan Kementerian Kesehatan Republik Indonesia [3], Indonesia berada pada urutan kelima dengan beban TB tertinggi di dunia. Estimasi prevalensi TB semua kasus adalah sebesar 660.000 kasus dan estimasi insidensi berjumlah 430.000 kasus baru per tahun. Jumlah kematian akibat TB diperkirakan 61.000 kematian per tahunnya. Untuk membuat diagnosis yang tepat, keberadaan mikroorganisme dalam dahak harus dibuktikan. Untuk membuktikan adanya *Mycobacterium tuberculosis* dalam dahak dibutuhkan media kultur sebagai baku emas untuk mendeteksi adanya *Mycobacterium tuberculosis*,

dimana *sample* dahak yang diperoleh dari pasien ditanam ke media ini selama 45 hari, setelah itu media kultur diperiksa untuk melihat tanda-tanda reproduksi bakteri [4]. Namun metode ini membutuhkan waktu yang lama dan biaya yang mahal.

Beberapa penelitian sebelumnya telah menggunakan teknik *machine learning* untuk memudahkan diagnosis tuberkulosis, diantaranya seperti *bagging*, *AdaBoost*, *Random forest tree*, ANFIS, PART, *Roughs sets* dan *data mining* [4][5][6]. Namun jika dilihat dari penelitian-penelitian yang sudah dilakukan tersebut sebagian besar menggunakan data medis dari rumah sakit setempat sehingga terdapat kemungkinan adanya perbedaan karakteristik data medis misalnya jumlah fitur yang digunakan.

Pada penelitian ini diusulkan metode seleksi *rule* (aturan) menggunakan *Rough Set Theory (RST)* untuk mendiagnosis penyakit tuberkulosis dengan data yang digunakan diperoleh dari rumah sakit umum Dr. Sardjito Yogyakarta. Metode ekstrakstraksi *rule* pada tahap awal diusulkan pada penelitian ini untuk mereduksi *rule* yang berjumlah besar. Dengan batasan nilai *support* terhadap sejumlah *rule* yang masih berjumlah besar, didapatkan *rule* yang lebih sedikit. Teknik RST kemudian digunakan untuk menyeleksi sejumlah *rule* hasil reduksi agar didapatkan *rule* yang lebih sedikit, dengan tetap menjaga kualitas performa klasifikasi dari sebelum dilakukan proses seleksi dan mempermudah dalam pengambilan keputusan.

2. Metode

Penelitian ini menggunakan metode seleksi *rule* untuk mendiagnosis penyakit tuberkulosis. Data yang digunakan diperoleh dari rumah sakit umum Dr. Sardjito Yogyakarta. Proses seleksi terbagi atas dua tahap, pertama ekstraksi *rule* diterapkan untuk mereduksi sejumlah *rule* yang berjumlah besar dengan batasan nilai *support*, kemudian seleksi *rule* dengan RST digunakan untuk menyeleksi *rule* hasil reduksi untuk mendapatkan *rule* yang lebih sedikit dengan tetap menjaga kualitas performa klasifikasi dari sebelum dilakukan proses seleksi.

2.1 Dataset Tuberkulosis

Dataset Tuberkulosis diperoleh dari data rekam medik pasien di rumah sakit umum Dr. Sardjito Yogyakarta. Proses pengambilan data dilakukan dengan membaca hasil rekam medik pasien yang pernah dirawat di rumah sakit dan didiagnosis penyakit tuberkulosis dengan bantuan dokter paru rumah sakit umum Dr. Sardjito. Data yang diperoleh sebanyak 183 objek dengan jumlah atribut sebanyak 23 dan satu atribut kelas. Atribut dari kumpulan data medis ini dikategorikan menjadi tiga kelompok, yaitu: (1) data demografi dan hasil temuan klinis, (2) hasil temuan laboratorium medis, dan (3) hasil temuan radiologi. Tabel 1 menunjukkan daftar atribut dari dataset tuberkulosis yang digunakan pada penelitian ini.

Tabel 1. Atribut tuberkulosis

Nama variabel	Nama variabel
1. Gender	13. Haemoptysis
2. Age group	14. Fever
3. Malaise	15. Erythrocyte
4. Arthralgia	16. Haematocrit
5. Exhaustion	17. Haemoglobin
6. Unwillingness for work	18. Leucocyte
7. Loss of appetite	19. Active specific lung lesion
8. Loss in weight	20. Calcific tissue
9. Sweating at nights	21. Cavity
10. Chest pain	22. Pneumonic infiltration
11. Back pain	23. Pleural effusion

Pada kelompok pertama Parameter *gender* menunjukkan apakah pasien seorang laki-laki atau perempuan. Parameter ini bernilai biner. Parameter *age group* menunjukkan kelompok usia pasien dengan nilai numerik. Parameter *Malaise*, *arthralgia*, *exhaustion*, *unwillingness for work*, *loss of appetite*, *loss in weight*, *sweating at nights*, *chest pain*, *back pain*, *coughing* dan *Haemoptysis* adalah parameter bernilai biner. Nilai ini menunjukkan nilai positif atau negatif pada pasien. *Fever* berarti pasien dalam keadaan demam, parameter ini diklasifikasikan menjadi 3 kategori, yaitu: " 0 " berarti demam biasa yang hampir 36,5 derajat Celcius, "1" berarti demam dalam rentang tinggi,

dan "2" berarti nilai demam *subfebrile* yang tidak melebihi 38,5 derajat Celcius.

Pada hasil temuan laboratorium medis, kelompok indikategorikan dalam beberapa parameter tes darah, diantaranya *Erythrocyte* adalah sel-sel darah merah. *Haematocrit* adalah rasio volume yang ditempati oleh sel darah merah untuk volume seluruh darah. *Haemoglobin* adalah *oxygen-transport metalloproteinase* yang mengandung besi dalam sel darah merah. *Leucocytes* adalah sel darah putih, parameter ini bertanggung jawab untuk mempertahankan tubuh terhadap infeksi, penyakit dll. Parameter hasil temuan laboratorium medis ini bernilai numerik.

Pada hasil temuan radiologi, parameter *active specific lung lesion* menunjukkan apakah ada bukti radiologis dari tuberkulosis paru lesi pada pasien. Parameter *Calcific tissue* menunjukkan apakah pasien memiliki penyakit tuberkulosis sebelumnya. Parameter *Cavity* menyatakan apakah ada pembukaan seperti luka pada paru-paru pasien atau tidak. Parameter *Pneumonic infiltration* menunjukkan jika ada radang paru-paru seperti luka yang terlihat di X-ray pada dada pasien, dan Parameter *Pleural effusion* berarti akumulasi cairan yang berlebihan pada selaput paru-paru. Parameter hasil temuan radiologi ini merupakan parameter bernilai biner.

2.2 Rough Set Theory (RST)

Rough set theory (RST) dikembangkan oleh Zdzislaw Pawlak pada tahun 1980-an [8], RST ini sangat berguna untuk menemukan hubungan dalam data yang disebut pengetahuan. Hasil penemuan pengetahuan berupa *rule* (aturan) yang mudah dimengerti dan bermakna, yang dihasilkan dari ekstraksi pola data. Metode RST muncul sebagai salah satu metode matematika untuk mengelola ketidakpastian, ambiguitas dan ketidakjelasan dari hubungan data yang tidak lengkap dan sulit dimodelkan secara matematis.

Untuk $S=(U,A)$ dan $B \subseteq A$, dimana $a \in B$, dengan :
 $INDs(B) = \{(x, x') \in U \times U \mid \forall a \in B, a(x) = a(x')\}$ dapat dikatakan bahwa a dapat diabaikan dalam B , dan :
 $INDs(B) = INDs(B - \{a\})$ jika sangat diperlukan a dapat diabaikan.

Himpunan B dikatakan independen jika semua atributnya diperlukan. Setiap subset B' dari B disebut *reduct* dari B jika B' adalah independen dan $INDs(B') = INDs(B)$.

Reduct dapat didefinisikan sebagai subset minimal dari beberapa atribut yang memiliki hasil klasifikasi yang sama. Dengan kata lain, atribut yang bukan unsur *reduct* merupakan *redundant* dari klasifikasi. *Reducts* relatif didasarkan pada elemen objek-objek tertentu

2.3 Rule Extraction dan Rule Selection

Ekstraksi *rule* (aturan) didasarkan pada nilai *support* dari masing-masing *rule*. Tujuannya untuk mereduksi sejumlah *rule* sehingga didapatkan jumlah *rule* yang lebih sedikit. Hal ini telah dilakukan pada penelitian sebelumnya [7]. Misalnya untuk,

$$DS = (U, C \cup D)$$

Yang merupakan tabel keputusan dari $\forall x \in U$, maka $c_1(x), \dots, c_k(x)$, $d(x)$ dapat didefinisikan, dimana $\{c_1, \dots, c_k\} = C$ dan $\{d\} = D$. *Decision rule* didapatkan dari,

$$c_1(x), \dots, c_2(x) \rightarrow d(x).$$

C merupakan *reduct* dari atribut kondisi yang merupakan hasil reduksi dari tabel keputusan.

Seleksi *rule* RST dilakukan karena *rule* yang didapat masih terlalu banyak dan panjang. Untuk menyerederhanakan jumlah *rule* dapat dilakukan melalui metode RST. Jika $R = \{Rule_1, Rule_2, \dots, Rule_j\}$ merupakan *subset rule* yang didapatkan dari *Rough Set* sebagai tabel keputusan yang baru, dimana *rule* berlaku sebagai subset atribut. Nilai 1 dari atribut $Rule_a$ jika objek x_b pada *decision* (d) mempunyai nilai yang sama dengan tabel keputusan dan bernilai 0 jika tidak bernilai sama. Nilai pada kolom (atribut) $j+1$ sama dengan nilai *decision*, dengan $a = 1, \dots, j$ dan $b = 1, \dots, i$. Tabel keputusan baru dapat direduksi menggunakan konsep *rough set*. *Reduct* yang didapat dari hasil reduksi atribut merupakan *rule* yang memiliki nilai akurasi yang besar. Untuk selanjutnya sejumlah *rule* hasil reduksi RST dapat diuji coba nilai akurasi nya menggunakan data uji kembali. Tabel 2 memperlihatkan *rule* dan nilai *support*, dan Tabel 3 memperlihatkan tabel keputusan dari seleksi *rule* yang akan dilakukan menggunakan RST.

Tabel 2. Contoh n *rule* dan nilai support

<i>rule</i>	<i>support</i>
$Rule_1$	47
$Rule_2$	46
$Rule_3$	45
$Rule_4$	45
$Rule_4$	44
.	.
.	.
.	.
$Rule_{n-3}$	2
$Rule_{n-2}$	1
$Rule_{n-1}$	1
$Rule_n$	1

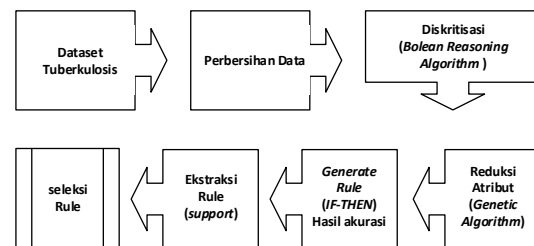
Tabel 3. Tabel keputusan dengan *rule* sebagai atribut

$x \in U$	$Rule_1$	$Rule_2$...	$Rule_{j-1}$	$Rule_j$	D
x_1	0	1	...	1	1	Yes
x_2	0	0	...	0	0	No

$x \in U$	$Rule_1$	$Rule_2$...	$Rule_{j-1}$	$Rule_j$	D
x_3	0	0	...	0	0	No
.
.
x_{i-2}	0	1	...	0	1	Yes
x_{i-1}	0	0	...	0	0	No
x_i	0	1	...	1	1	Yes

2.4 Langkah-langkah penelitian

Pada tahapan penelitian, pertama dataset tuberkulosis dilakukan pembersihan data dengan membuang data yang terdapat *missing value* yang terlalu banyak. Selanjutnya data yang bernilai numerik didiskritisasi dengan algoritme *boolean reasoning*. Data hasil diskritisasi tersebut kemudian dilakukan reduksi atribut menggunakan *Genetic algorithm* dengan pilihan *object related reduct* yang sekaligus menghasilkan *rule*. Untuk proses selanjutnya dilakukan proses ekstraksi dan seleksi *rule* menggunakan teknik RST. Penelitian ini dilakukan dengan menggunakan perangkat lunak Rosetta Rough Set Toolkit. Langkah-langkah pada penelitian ini dapat dilihat pada gambar 1 dibawah ini.



Gambar 1. Langkah-langkah penelitian

3. Hasil dan Pembahasan

3.1 Data Preprocessing

Pada proses awal dataset dilakukan pembersihan data dengan membuang data yang tidak lengkap atau *missing value* dan dibuat menjadi dua kelas yaitu kelas positif diberi label ‘yes’ dan kelas negatif diberi label ‘no’. Untuk kelas positif merupakan data pasien yang menderita penyakit tuberkulosis dan kelas negatif adalah data pasien yang tidak menderita penyakit tuberkulosis, dengan rincian kelas positif berjumlah 103 objek dan kelas negatif berjumlah 80 objek. Data hasil *preprocessing* dapat dilihat pada Tabel 4.

Tabel 4. Data hasil *preprocessing*

No	Gender	Age	.	cavity	pleural effusion	Status
1	1	58	.	0	0	yes
2	0	55	.	0	0	yes

No	Gender	Age	cavity	pleural effusion	Status
3	1	40	0	0	yes
4	1	1	0	0	yes
5	0	43	0	0	yes
6	0	34	0	0	yes
7	1	42	0	0	no
177	0	5	0	0	no
178	1	19	0	0	yes
179	1	25	0	0	no
180	1	57	0	0	yes
181	1	15	0	0	no
182	0	4	0	0	no
183	1	46	0	0	no

Proses selanjutnya data yang bernilai 100numerik dilakukan proses diskretisasi dengan menggunakan *boolean reasoning*. Pada penelitian ini data yang bernilai 100numerik terdapat pada atribut *age*, *fever*, *erythrocyte*, *haematocrit*, *haemoglobin* dan *leucocyte*. Hasil diskretisasi dapat dilihat pada Tabel 5, dimana nilai $[*,12)$ berarti $Haemoglobin < 12$, $[12,13)$ berarti $12 \leq Haemoglobin < 13$, dan $[13,*)$ berarti $Haemoglobin \geq 13$.

Tabel 5. Hasil diskretisasi

Atribut numerik	Nilai diskret
Age	$[*,16)$, $[16,23)$, $[23, 30)$, $[30,43)$, $[43, 54)$, $[54, 66)$, $[66, *)$
Fever	$[*, 35.8)$, $[35.8, 36.2)$, $[36.2, *)$
Erythrocyte	$[*, 4.1)$, $[4.1, 4.4)$, $[4.4, 4.7)$, $[4.7, *)$
Haematocrit	$[*, 36)$, $[36, *)$
Haemoglobin	$[*, 12)$, $[12, 13)$, $[13, *)$
Leucocyte	$[*, 5250)$, $[5250, 6250)$, $[6250, *)$

Dapat dilihat pada Tabel 5 bahwa proses diskretisasi membagi fitur-fitur dengan tipe data numerik menjadi tipe nominal dengan interval tertentu. Fitur *age* didiskretisasi menjadi 7 interval, fitur *Fever* didiskretisasi menjadi 3 interval, fitur *Erythrocyte* didiskretisasi menjadi 4 interval, fitur *Haematocrit* didiskretisasi menjadi 2 interval, fitur *Haemoglobin* didiskretisasi menjadi 3 interval, dan fitur *Leucocyte* didiskretisasi menjadi 3 interval.

Selanjutnya data di-split menjadi menjadi 50:50, dengan 92 objek data dijadikan data latih dan 91 objek data dijadikan data uji, dan selanjutnya data di-reduct menggunakan *ROSETTA Genetic algorithm* dan *discernibility object related*, didapatkan minimal subset atribut tereduksi, dan selanjutnya akan dilakukan *generate rule*, kemudian *rule* yang didapat diujikan ke data uji untuk dilihat hasil klasifikasinya. Dari proses *reduct* yang dilakukan diperoleh jumlah *reduct* sebanyak 622, jumlah *rule* sebanyak 1303

dengan akurasi 0.956 atau 95.6%. Pada Tabel 6 diperlihatkan *rule* yang diperoleh dari proses *reduct*.

Tabel 6. Rule dari proses *reduct*

No	Decision Rule	Support
1	unwillingness_for_work(1) AND loss_of_appetite(1) => status(yes)	47
2	exhaustion(1) AND loss_of_appetite(1) => status(yes)	46
3	unwillingness_for_work(1) AND loss_in_weight(1) => status(yes)	45
4	exhaustion(1) AND loss_in_weight(1) => status(yes)	44
5	malaise(1) AND exhaustion(1) AND loss_in_weight(1) => status(yes)	44
1299	age([23, 30)) AND erythrocyte([4.7, *) AND haematocrit([*, 36)) => status(no)	1
1300	age([23, 30)) AND loss_in_weight(0) AND erythrocyte([4.7, *) => status(yes)	1
1301	fever([*, 35.8)) AND haemoglobin([*, 12)) AND leucocyte([5250, 6250)) => status(no)	1
1302	age([66, *) AND erythrocyte([4.7, *) AND haematocrit([36, *) => status(yes)	1
1303	exhaustion(1) AND coughing(0) AND erythrocyte([4.4, 4.7)) => status(no)	1

3.2 Rule Extraction dan Rule Selection

Pada proses ini *rule* yang didapatkan kemudian diekstraksi menggunakan *basic filtering*, dimana parameter yang digunakan adalah berdasar nilai *support*. Pada penelitian ini digunakan nilai *support*=22, yang artinya *rule* yang memiliki nilai *support* dibawah 22 akan dihapus. Hasil ekstraksi *rule* dapat dilihat pada Tabel 7.

Tabel 7. Rule terekstraksi

No	Decision Rule	Support
1	unwillingness_for_work(1) AND loss_of_appetite(1) => status(yes)	47
2	exhaustion(1) AND loss_of_appetite(1) => status(yes)	46
3	unwillingness_for_work(1) AND loss_in_weight(1) => status(yes)	45
4	malaise(1) AND exhaustion(1) AND loss_in_weight(1) => status(yes)	44
5	exhaustion(1) AND loss_in_weight(1) => status(yes)	44
6	malaise(1) AND exhaustion(1) AND chest_pain(1) => status(yes)	40
7	malaise(1) AND unwillingness_for_work(1) AND chest_pain(1) => status(yes)	40
8	loss_of_appetite(1) AND chest_pain(1) => status(yes)	35
9	gender(1) AND malaise(1) AND unwillingness_for_work(1) => status(yes)	33
10	loss_in_weight(1) AND chest_pain(1) => status(yes)	33
11	exhaustion(0) AND chest_pain(0) AND pneumonic_infiltration(0) => status(no)	33
12	gender(1) AND malaise(1) AND exhaustion(1) => status(yes)	32
13	unwillingness_for_work(0) AND chest_pain(0) AND pneumonic_infiltration(0) => status(no)	32
14	malaise(1) AND exhaustion(1) AND coughing(1) => status(yes)	32
15	unwillingness_for_work(0) AND chest_pain(0) AND back_pain(0) AND pleural_effusion(0) => status(no)	32
16	malaise(1) AND unwillingness_for_work(1) AND coughing(1) => status(yes)	32

No	Decision Rule	Support
17	unwillingness_for_work(1) AND coughing(1) => status(yes)	32
18	unwillingness_for_work(0) AND chest_pain(0) AND pleural_effusion(0) => status(no)	32
19	exhaustion(1) AND coughing(1) => status(yes)	32
20	loss_of_appetite(1) AND loss_in_weight(1) AND chest_pain(1) => status(yes)	32
21	loss_in_weight(0) AND chest_pain(0) AND pneumonic_infiltration(0) => status(no)	31
22	loss_in_weight(0) AND chest_pain(0) AND active_specipic_lung_lesion(0) AND pneumonic_infiltration(0) => status(no)	31
23	malaise(1) AND loss_of_appetite(1) AND coughing(1) => status(yes)	30
24	loss_of_appetite(1) AND coughing(1) => status(yes)	30
25	malaise(0) => status(no)	29
26	loss_of_appetite(0) AND chest_pain(0) AND pneumonic_infiltration(0) => status(no)	29
27	loss_in_weight(1) AND coughing(1) => status(yes)	28
28	malaise(1) AND unwillingness_for_work(1) AND leucocyte([6250, *]) => status(yes)	27
29	unwillingness_for_work(1) AND leucocyte([6250, *]) => status(yes)	27
30	unwillingness_for_work(1) AND haematocrit([*, 36]) => status(yes)	27
31	gender(1) AND unwillingness_for_work(1) AND loss_in_weight(1) => status(yes)	27
32	exhaustion(1) AND haematocrit([*, 36]) => status(yes)	27
33	exhaustion(1) AND leucocyte([6250, *]) => status(yes)	26
34	exhaustion(1) AND unwillingness_for_work(1) AND leucocyte([6250, *]) => status(yes)	26
35	malaise(1) AND exhaustion(1) AND leucocyte([6250, *]) => status(yes)	26
36	malaise(1) AND loss_of_appetite(1) AND haematocrit([*, 36]) => status(yes)	25
37	sweating_at_nights(0) AND coughing(0) AND pneumonic_infiltration(0) => status(no)	24
38	back_pain(0) AND coughing(0) AND active_specipic_lung_lesion(0) AND pleural_effusion(0) => status(no)	24
39	malaise(1) AND loss_in_weight(1) AND haematocrit([*, 36]) => status(yes)	24
40	back_pain(0) AND coughing(0) AND pneumonic_infiltration(0) => status(no)	24
41	pneumonic_infiltration(1) => status(yes)	23
42	chest_pain(0) AND coughing(0) AND active_specipic_lung_lesion(0) AND pleural_effusion(0) => status(no)	23
43	chest_pain(0) AND back_pain(0) AND coughing(0) AND pneumonic_infiltration(0) => status(no)	23
44	unwillingness_for_work(1) AND fever([36.2, *]) => status(yes)	23
45	exhaustion(1) AND fever([36.2, *]) => status(yes)	23
46	exhaustion(1) AND chest_pain(1) AND coughing(1) => status(yes)	23
47	chest_pain(0) AND coughing(0) AND pneumonic_infiltration(0) => status(no)	23
48	exhaustion(0) AND coughing(0) AND pneumonic_infiltration(0) => status(no)	22
49	malaise(1) AND loss_of_appetite(1) AND fever([36.2, *]) => status(yes)	22
50	loss_in_weight(0) AND coughing(0) AND pleural_effusion(0) => status(no)	22
51	loss_in_weight(0) AND coughing(0) AND pneumonic_infiltration(0) => status(no)	22
52	loss_in_weight(0) AND coughing(0) AND haemoptysis(0) AND	22

No	Decision Rule	Support
	pleural_effusion(0) => status(no)	
53	malaise(1) AND loss_in_weight(1) AND fever([36.2, *]) => status(yes)	22

Dapat dilihat pada Tabel 6 bahwa hasil ekstraksi *ruled* dengan nilai *support* =22 menghasilkan jumlah *rule* sebanyak 53 dengan nilai akurasi tetap sebesar 0.956 atau 95.6%.

Selanjutnya, dengan teknik seleksi *rule* berbasis RST dibuat tabel keputusan baru dari *rule* yang sudah terekstraksi. *Rule* hasil ekstraksi disusun menjadi atribut dari tabel keputusan baru, dan 92 objek dari data latih digunakan sebagai *instance*. Kemudian data di-*reduct* menggunakan ROSETTA Johnson's algorithm dan *discernibilityfull object* diperoleh lima *reduct* yang merepresentasikan lima *rule* terseleksi. Kelima *reduct* tersebut adalah *Rule1*, *Rule6*, *Rule11*, *Rule37*, dan *Rule41*. *Rule* terseleksi dapat dilihat pada Tabel 8.

Tabel 8. *Rule* hasil seleksi

No	Decision Rule	Rule No.
1	unwillingness_for_work(1) AND loss_of_appetite(1) => status(yes)	1
2	malaise(1) AND exhaustion(1) AND chest_pain(1) => status(yes)	6
3	exhaustion(0) AND chest_pain(0) AND pneumonic_infiltration(0) => status(no)	11
4	sweating_at_nights(0) AND coughing(0) AND pneumonic_infiltration(0) => status(no)	37
5	pneumonic_infiltration(1) => status(yes)	41

Proses selanjutnya dilakukan pengujian *rule* terseleksi terhadap 91 data uji. Dari 5 *rule* terseleksi didapat hasil akurasi sebesar 0.956 atau 95.6%. Nilai akurasi ini sama jika dibandingkan dengan nilai akurasi pada saat sebelum dilakukan proses seleksi dengan RST, yang artinya dengan jumlah *rule* yang lebih sedikit kualitas performaklasifikasinya tetap sama.

4. Kesimpulan

Dari hasil percobaan yang telah dilakukan, dapat disimpulkan bahwa:

- Metode seleksi *rule* berbasis RST telah berhasil digunakan untuk menyeleksi lima *rule* dari 1303 *rule* tanpa mengurangi kualitas nilai akurasi yaitu sebesar 95.6%.
- *Rule* terseleksi dapat digunakan sebagai basis pengetahuan untuk sistem pendukung pengambilan keputusan dalam mendiagnosis penyakit tuberkulosis.
- Untuk penelitian selanjutnya diperlukan lebih banyak data untuk meningkatkan proses pembuatan model. Optimasi model diperlukan

untuk meningkatkan akurasi dan pengambilan keputusan yang cepat.

Ucapan Terima Kasih

Dalam melakukan penelitian ini, penulis telah mendapatkan banyak dukungan dan bantuan dari berbagai pihak. Penulis mengucapkan terimakasih yang tak terhingga kepada:

1. Bapak Noor Akhmad Setiawan, S.T., M.T., Ph.D selaku dosen pembimbing utama, dan Ibu Indriana Hidayah, S.T., M.T. selaku dosen pembimbing pendamping, yang telah dengan penuh kesabaran dan ketulusan memberikan ilmu dan bimbingan terbaik kepada penulis.
2. Para Bapak/Ibu Dosen Program Studi S2 Teknik Elektro Universitas GadjahMada yang telah memberikan bekal ilmu kepada penulis.
3. Para Bapak/Ibu Karyawan/wati Program Studi S2 Teknik Elektro Universitas Gadjah Mada yang telah membantu penulis dalam proses belajar.

Daftar Pustaka

- [1] Departemen Kesehatan Republik Indonesia, "Pedoman Nasional Penanggulangan Tuberkulosis", edisi kedua. Cetakan pertama. 2006.
- [2] Perhimpunan Dokter Paru Indonesia, "Tuberkulosis, Pedoman Diagnosis dan Penatalaksanaan di Indonesia", 2006.
- [3] Kementerian Kesehatan Republik Indonesia Direktorat Jendral Pengendalian Penyakit dan Penyehatan Lingkungan, "Terobosan menuju akses universal Strategi Nasional Pengendalian TB di Indonesia 2010-2014", 2011
- [4] Tamer Uçar, Adem Karahoca, Dilek Karahoca, "Tuberculosis disease diagnosis by using adaptive neuro fuzzy inference system and rough sets", *Neural Computing and Applications Appl.*, Vol 23, no. 2, pp. 471-483 2013
- [5] Asha.T, Dr. S. Natarajan, Dr. K.N.B. Murthy, "Diagnosis of Tuberculosis using Ensemble methods", *Computer Science and Information Technology (ICCSIT)*, International Conference, Vol.8 , pp: 409 – 412, 2010
- [6] Tamer Uçar, Adem Karahoca, "Predicting existence of Mycobacterium tuberculosis on patients using data mining approaches," *Procedia Computer Science*, Vol 3, Pages 1404-1411, 2011
- [7] N.A.Setiawan,P.A.Venkathachalam,and Ahmad Fadzil M.H,"Rule Selection for Coronary Artery Disease Diagnosis Based

on Rough Set, *International Journal of Recent Trends Engineering*, 2009

- [8] Z. Pawlak, "Rough Sets," *International Journal of Computer and Information Sciences*, vol. 11, pp. 341-355, 1982