

# IMPLEMENTASI TEKNIK DATA CLEANING DAN TEKNIK ROUGHSET PADA DATA TIDAK LENGKAP DALAM DATA MINING

Jasmir, S.Kom, M.Kom

Dosen Teknik Informatika STIKOM Dinamika Bangsa Jambi

Email : ijay\_jasmir@yahoo.com

## Abstrak

Penelitian ini dilakukan untuk mempelajari teknik *Data Cleaning* yang diimplementasikan untuk pencarian *Knowledge* terhadap data tidak lengkap dalam data mining. Selanjutnya data dianalisa untuk mempelajari pola dari masing-masing teknik yang digunakan dari *data cleaning* yaitu *mean substitusi* dan *mean interval*. Berdasarkan pola tersebut, teknik yang lebih baik dipilih untuk pembangunan sebuah sistem yang diharapkan bisa dimanfaatkan oleh manajemen dalam membantu proses pengambilan keputusan. Pengambilan keputusan dilakukan dengan memanfaatkan teknik *roughset*. Dari eksperimen yang dilakukan diketahui bahwa masing-masing teknik tersebut mempunyai kelebihan dan kekurangan. Akhir dari penelitian adalah ditemukannya *knowledge* dari basis data tidak lengkap untuk pengambilan keputusan yang tepat walau data yang tersedia tidak selalu lengkap.

Kata Kunci : *Data Cleaning, Roughset, Data Mining, Data Tidak Lengkap*

## 1. Pendahuluan

Beberapa tahun belakangan, penggunaan basis data semakin meningkat cepat dan jumlah data yang ditangani pun semakin besar jumlahnya. Dengan semakin besarnya ukuran data tersebut, proses ekstraksi informasi dengan cara analisa secara manual semakin sulit, rentan kesalahan, memakan waktu lama, dan mengandung subjektifitas tinggi. Karena itu timbul tuntutan untuk memunculkan cara dalam melakukan penarikan informasi secara cepat dan akurat yang dapat menangani pengolahan data ukuran besar. Hal inilah yang mendorong timbulnya *Knowledge Discovery in Databases* (KDD). KDD pada dasarnya adalah serangkaian kegiatan yang melakukan transformasi *low-level data* menjadi suatu bentuk informasi yang mudah dicerna oleh manusia, contohnya berupa laporan.

Salah satu proses yang terlibat dalam KDD adalah *Data Mining* yaitu sebuah langkah dalam proses KDD yang melakukan analisis terhadap data dan menemukan algoritma yang menghasilkan sejumlah pola-pola tertentu dari data tersebut. Tujuan utama dari *Data Mining* adalah melakukan prediksi nilai dari variabel-variabel tertentu, serta mendeskripsikan data-data ke dalam bentuk representasi yang mudah dimengerti oleh manusia. Salah satu pemodelan yang dihasilkan dari proses *Data Mining* adalah teknik-teknik *Artificial Intelligence* (AI) untuk proses pencarian *Knowledge*. Salah satu teknik AI yang digunakan adalah teknik *Rough Set*.

## 2. Data Mining

*Data mining* merupakan serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data. Definisi lain data mining adalah sebagai proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar. *Data mining* juga diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Istilah data mining kadang disebut juga *knowledge discovery*. Istilah data mining dan *Knowledge Discovery in Database* (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar [2]

*Data Mining* adalah proses inti dari rangkaian proses dalam KDD (*Knowledge Discovery in Databases*). Seperti yang telah dijelaskan dalam sub-bab sebelumnya, *Data Mining* adalah sebuah proses pencarian pola dari data dan merepresentasikannya kedalam bentuk yang lebih mudah dipahami.

## 3. Metode Rought Set

*Rough set* adalah sebuah teknik matematika yang dikembangkan oleh Pawlack pada tahun 1980. *Rough Set* salah satu teknik *data mining* yang digunakan untuk menangani masalah *Uncertainty, Imprecision* dan *Vagueness* dalam aplikasi *Artificial*

*Intelligence (AI)*. *Rough set* merupakan teknik yang efisien untuk *Knowledge Discovery in Database (KDD)* dalam tahapan proses dan *Data Mining*. [10]

*Rough Set* merupakan teknik yang efisien untuk *Knowledge Discovery* dalam Database (KDD) proses dan *Data Mining*. Secara umum, teori *Rough Set* telah digunakan dalam banyak aplikasi seperti *medicine, pharmacology, business, banking, engineering design, image processing* dan *decision analysis*. *Rough Set* merupakan teknik yang efisien untuk KDD proses dan *Data Mining*.

#### 4. Data Cleaning

*Teknik data cleaning* adalah suatu teknik yang digunakan untuk menangani data yang tidak lengkap. Proses *data cleaning* ini juga mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak. [6]

*Data Cleaning* juga Melakukan proses *enrichment*, yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi *eksternal*

### 5. Pembahasan

#### 5.1. Teknik Data Cleaning

Dengan menggunakan proses data *cleaning* dengan teknik *mean substitusi* dijelaskan, data yang tidak lengkap diisi dengan nilai rata-rata dari sampel algoritma dari teknik *mean substitusi*.

#### 5.2. Studi Kasus

Untuk melakukan uji coba pencarian knowledge dalam *datamining* disediakan data sebanyak 20 record sebagai sampel. Sampel data tersebut adalah untuk membantu dalam pengambilan keputusan nantinya, yaitu data nasabah pelanggan Perusahaan Daerah Air Minum (PDAM) Kota Sungai Penuh. Dimana dalam analisa ini nantinya output yang diharapkan adalah kenaikan tarif pelanggan PDAM. Kenaikan tersebut dianalisa tidak bersifat *komprensif* melainkan kenaikan tarif tersebut secara proporsional.

##### 5.2.1 Analisa Data

Untuk mendapatkan *output* yang diinginkan perlu adanya variabel-variabel *input*. Dimana pada data diatas yang menjadi variabel-variabel *input* adalah Jenis Bangunan, Luas bangunan, dan letak bangunan,

sedangkan variabel outputnya adalah variabel Tarif. Masing-masing variabel input dan variabel output yang digunakan akan dijelaskan sebagai berikut:

**Tabel 3-1 Variabel Input Output**

##### A. Jenis Bangunan

Nilai	Kategori
1	Mewah
2	Permanen
3	Semi Permanen

##### B. Luas Bangunan

Nilai	Kategori
1	<= 60 m2
2	60 m2 s/d 100 m2
3	100m2 s/d 200m2
4	>= 200 m2

##### C. Letak Bangunan

Nilai	Kategori
1	Jalan Utama
2	Jalan Lingkungan
3	Gang

##### D. Tarif

Nilai	Kategori
1	Kenaikan Tarif 15%
2	Kenaikan Tarif 10 %
3	Kenaikan Tarif 5 %

#### 5.2.2 Proses Teknik Data Cleaning

Berdasarkan data yang disajikan diatas maka terlihat ada data-data yang tidak lengkap yang disajikan dalam tabel 3-2 dibawah ini :

**Tabel 3-2 Data Tidak Lengkap**

Nama	A	B	C
Suhaeni	2	3	2
Atmajar Idris	2	3	1
Iskandar	2	2	1
Nasrul Effendi	?	3	1
Ellya	2	2	1

Edwar	2	3	1
Mardison Garogo	2	3	1
Zulfahmi	2	3	?
Emidarwita	2	2	1
Susilawati	3	?	1
Bilal Alfayyadh	2	2	2
Zayyan Athaya	2	?	1
Daris Alfajri	2	3	1
Siti Khadijah	2	2	1
Maulida	2	2	2
Lisna Wita	2	2	2
Fitri Daryanti	?	2	2
Dayang Antari	1	2	2
Dara Yustika	2	2	1
Sapta Putra	2	2	1
Hendra CB	2	2	?
Santoso	2	?	2
Imam Alif	2	2	2
Lisa Novita	2	2	2

Data dari tabel tidak lengkap diatas dilengkapi dengan teknik Data *Cleaning* seperti yang telah dijelaskan sebelumnya. Dalam kasus ini akan digunakan 2 buah teknik data *cleaning* untuk membandingkan algoritma teknik data *cleaning* mana yang *rulesnya* mendekati benar. Teknik Data *Clening* yang digunakan adalah teknik data *Mean Substitusi* dan teknik *Mean Interval*.

1) Algoritma Teknik Data Mean Substitusi

a) Jenis Bangunan =

$$\bullet (2+2+2+2+2+2+2+3+2+2+2+2+2+1+2+2+2+2+2)/24 = 48/24 = 2$$

b) Luas Bangunan =

$$\bullet (3+3+2+2+3+2+3+3+3+2+2+3+2+2+2+2+2+2+2+2)/24 = 2.6 = 3$$

c) Letak Bangunan =

$$\bullet (2+1+1+1+1+1+1+1+1+2+1+1+1+2+2+2+2+1+1+2+2+2)/24 = 1.2 = 1$$

**Tabel 3-3 Data Lengkap**

Nama	A	B	C
Suhaeni	2	3	2
Atmajar Idris	2	3	1

Iskandar	2	2	1
Nasrul Effendi	<b>2</b>	3	1
Ellya	2	2	1
Edwar	2	3	1
Mardison Garogo	2	3	1
Zulfahmi	2	3	<b>1</b>
Emidarwita	2	2	1
Susilawati	3	<b>3</b>	1
Bilal Alfayyadh	2	2	2
Zayyan Athaya	2	<b>3</b>	1
Daris Alfajri	2	3	1
Siti Khadijah	2	2	1
Maulida	2	2	2
Lisna Wita	2	2	2
Fitri Daryanti	<b>2</b>	2	2
Dayang Antari	1	2	2
Dara Yustika	2	2	1
Sapta Putra	2	2	1
Hendra CB	2	2	<b>1</b>
Santoso	2	<b>3</b>	2
Imam Alif	2	2	2
Lisa Novita	2	2	2

1) Algoritma Teknik Mean Interval

a) Jenis Bangunan

$$\bullet \text{Rata-rata awal} = (2+2+2+2+2+2+2+2)/8 = 16/8 = 2$$

$$\bullet \text{Rata-rata tengah} = (2+3+2+2+2+2+2+2)/8 = 17/8 = 2$$

$$\bullet \text{Rata-rata akhir} = (2+1+2+2+2+2+2+2)/8 = 15/8 = 2$$

2) Luas Bangunan

$$\bullet \text{Rata-rata awal} = (3+3+2+2+3+2+3+3)/8 = 19/8 = 2$$

$$\bullet \text{Rata-rata tengah} = (3+3+2+2+2+3+2+2)/8 = 19/8 = 2$$

$$\bullet \text{Rata-rata akhir} = (2+2+2+2+2+2+2+2)/8 = 16/8 = 2$$

3) Letak Bangunan

$$\bullet \text{Rata-rata awal} = (2+1+1+1+1+1+1+1)/8 = 9/8 = 1$$

$$\bullet \text{Rata-rata tengah} = (1+2+1+1+1+2+2+2)/8 = 12/8 = 1.5 = 2$$

$$\bullet \text{Rata-rata akhir} = (2+2+2+1+1+2+2+2)/8 = 14/8 = 1.75 = 2$$

**Tabel 3-4 Data Lengkap**

Nama Pelanggan	A	B	C
Suhaeni	2	3	2
Atmajar Idris	2	3	1
Iskandar	2	2	1
Nasrul Effendi	2	3	1
Ellya	2	2	1
Edwar	2	3	1
Mardison Garogo	2	3	1
Zulfahmi	2	3	1
Emidarwita	2	2	1
Susilawati	3	2	1
Bilal Alfayyadh	2	2	2
Zayyan Athaya	2	2	1
Daris Alfajri	2	3	1
Siti Khadijah	2	2	1
Maulida	2	2	2
Lisna Wita	2	2	2
Fitri Daryanti	2	2	2
Dayang Antari	1	2	2
Dara Yustika	2	2	1
Sapta Putra	2	2	1
Hendra CB	2	2	2
Santoso	2	2	2
Imam Alif	2	2	2
Lisa Novita	2	2	2

Dari dua algoritma teknik data *cleaning* diatas ternyata terjadi perbedaan nilai data yaitu pada variabel luas bangunan, dimana dengan metode algoritma *mean substitusi* data tidak lengkap diisi 3 setelah dihitung sedangkan dengan menggunakan algoritma teknik *mean interval* data tidak lengkap diisi dengan 2 setelah dijalankan algoritmanya. Perbedaan juga terlihat pada variabel Letak Bangunan pada interval nilai rata-rata akhir dimana dengan algoritma teknik *Mean Substitusi* bernilai 1 sedangkan dengan algoritma teknik *Mean Interval* bernilai 2. Perbedaan inilah nantinya yang akan diuji mana yang lebih mendekati fakta setelah diekstrak dari data mining dalam metode *Rough Set*.

### 5.2.3 Proses Data Mining

Untuk mendapatkan *reduct* dan *rule* perlu dilakukan *ekstraksi* dari database untuk mendapatkan *Knowledge*, selanjutnya *knowledge* tersebut digunakan untuk membuat *decision* (pengambilan keputusan)

*Database* yang akan diekstrak berisi kumpulan data yang berasal dari pengalaman-pengalaman masa lalu seorang pakar tentang pemecahan masalah, disimpan dalam database dan digunakan kembali untuk pemecahan masalah berikutnya. Salah satu *ekstraksi Database* adalah dengan menggunakan *Case Based Reasoning (CBR)* seperti yang sudah dijelaskan pada bab II. Dalam penelitian ini untuk mengekstrak *database* penulis menggunakan CBR

### Langkah-langkah Ekstraksi Data dalam Data Mining pada Rough Set

#### 1. Decision System

$$DS = \{U, (A,C)\}$$

Dimana : A = Conditional Attribute  
U = Object  
C = Decision System

Dalam kasus diatas :

$$A = \{E1, E2..E24\}$$

$$U = \{\text{Jenis Bangunan, Luas}$$

Bangunan, Letak Bangunan}

$$C = \{\text{Tarif}\}$$

#### 2. Equivalence Class

*Equivalence Class* adalah Sekumpulan object yang akan dikelompokkan yang punya nilai atribut yang sama. Berikut ini tabel 3-5 memperlihatkan numerical representasi dari *equivalence class*

**Tabel 3-5 Equivalence Class dari Algoritma Teknik Mean Substitusi**

Class	A	B	C	D	Num_Obj
EC1	2	3	2	1	2
EC2	2	3	1	1	7
EC3	2	2	1	2	8
EC4	3	3	1	3	1
EC5	2	2	2	3	5
EC6	1	2	2	2	1

**Tabel 3-6 Equivalence Class dari Algoritma Teknik Mean Interval**

Class	A	B	C	D	Num_Obj
EC1	2	3	2	1	1
EC2	2	3	1	1	6
EC3	2	2	1	2	7
EC4	2	2	2	3	8
EC5	3	2	1	3	1
EC6	1	2	2	2	1

Dari kedua tabel equivalence diatas jelas sekali terlihat perbedaan yang ada pada Atribut Kondisi karena menggunakan 2 algoritma yang berbeda pada data cleaning. Perbedaan ini akan diteruskan pengujiannya pada langkah-langkah selanjutnya sehingga nantinya tentu saja akan menghasilkan *Reduct* yang berbeda pula yang akan menghasilkan *rules* yang berbeda pula.

### 3. Discernibility Matrix

*Discernibility Matrix* adalah Pengelompokan sejumlah *attribute* dimana yang dikelompokkan hanyalah *Attribute Conditional* saja.

#### A. Discernibility Matrix

Pada tabel 3-7 dibawah ini akan menjelaskan hasil dari *Discernibility Matrix* dari *Equivalence Class* dengan menggunakan metode algoritma *Mean Substitusi*.

**Tabel 3-7 Discernibility Matrix**

	EC1	EC2	EC3	EC4	EC5	EC6
EC1	X	C	BC	AC	B	AB
EC2	C	X	B	A	BC	ABC
EC3	BC	B	X	AB	C	AC
EC4	AC	A	AB	X	ABC	ABC
EC5	B	BC	C	ABC	X	A
EC6	AB	ABC	AC	ABC	A	X

Pada tabel 3-8 dibawah ini akan menjelaskan hasil dari *Discernibility Matrix* dari *Equivalence Class* dengan menggunakan metode algoritma *Mean Interval*.

**Tabel 3-8 Discernibility Matrix**

	EC1	EC2	EC3	EC4	EC5	EC6
EC1	X	C	BC	B	ABC	AB
EC2	C	X	B	BC	AB	ABC
EC3	BC	B	X	C	A	AB
EC4	B	BC	C	X	AC	A
EC5	ABC	AB	A	AC	X	AC
EC6	AB	ABC	AB	A	AC	X

### 4. Discernibility Matrix Modulo D

*Discernibility Matrix Modulo D* adalah Pengelompokan sejumlah *attribute* yang berbeda *attribute condisinya* dan berbeda pula *decisionnya*. Juga akan dijelaskan perbedaan *Discernibility Matrix Modulo D* dengan menggunakan dua algoritma data *cleaning* yang berbeda.. Pada tabel 3-9 dibawah ini akan menjelaskan hasil dari *Discernibility Matrix Modulo D* dari *Equivalence Class* dengan menggunakan metode algoritma *Mean Substitusi*.

**Tabel 3-9 Discernibility Matrix Modulo D**

	EC1	EC2	EC3	EC4	EC5	EC6
EC1	X	X	BC	AC	B	AB
EC2	X	X	B	A	BC	ABC
EC3	BC	B	X	AB	C	X
EC4	AC	A	AB	X	X	ABC
EC5	B	BC	C	X	X	A
EC6	AB	ABC	X	ABC	A	X

Pada tabel 3-10 dibawah ini akan menjelaskan hasil dari *Discernibility Matrix Modulo D* dari *Equivalence Class* dengan menggunakan metode algoritma *Mean Interval*.

**Tabel 3-10 Discernibility Matrix Modulo D**

	EC1	EC2	EC3	EC4	EC5	EC6
EC1	X	X	BC	B	ABC	AB
EC2	X	X	B	BC	AB	ABC
EC3	BC	B	X	C	A	X
EC4	B	BC	C	X	X	A
EC5	ABC	AB	A	X	X	AB
EC6	AB	ABC	X	A	AB	X

### 5. Reduct Calculation

*Reduct* adalah penyeleksian *attribute* minimal dari sekumpulan *attribute* kondisi yang menggunakan *Prime Implication Fungsi Boolean*. Kumpulan dari semua *Prime Implicant* mendeterminasi *set of reduct*. *Reduct* didapat dengan *CNF of Boolean Function*. Persamaan *Boolean* untuk *Mean Substitusi* sbb

**Tabel 3-11 Reduct**

Class	CNF of Boolean Fuction	Reduct
E1	$(BUC) \wedge (AUC) \wedge B \wedge (AUB)$	{a,b},{b,c}
E2	$B \wedge A \wedge (BUC) \wedge (AUBUC)$	{a,b}
E3	$(BUC) \wedge B \wedge (AUB) \wedge C$	{b,c}
E4	$(AUC) \wedge A \wedge (AUB) \wedge AUBUC$	{a}
E5	$B \wedge (BUC) \wedge C \wedge A$	{a,b,c}
E6	$(AUB) \wedge (AUBUC) \wedge (AUBUC) \wedge A$	{a}

Persamaan *Boolean* untuk *Mean Interval* sbb

**Tabel 3-12 Reduct**

Class	CNF of Boolean Fuction	Reduct
E1	$(BUC) \wedge B \wedge (AUBUC) \wedge (AUB)$	{a,b}, {b}
E2	$B \wedge (BUC) \wedge (AUB) \wedge (AUBUC)$	{a,b}, {b}
E3	$(BUC) \wedge B \wedge C \wedge A$	{a,b,c}
E4	$B \wedge BC \wedge C \wedge A$	{a,b,c}
E5	$(AUBUC) \wedge (AUB) \wedge A \wedge (AUB)$	{a}
E6	$(AUB) \wedge (AUBUC) \wedge A \wedge (AUB)$	{a}

## 6. Pencarian Knowledge ( Generating Rules)

*Knowledge ( Rules)* dari *ekstraksi data mining* yang terbentuk berdasarkan *equivalence class* yang terbentuk untuk metode algoritma *Mean Substitusi* adalah sebagai berikut :

**Tabel 3-13 Generating Rules**

Class	A	B	C	D	Reduct
EC1	2	3	2	1	{a,b}, {b,c}
EC2	2	3	1	1	{a,b}
EC3	2	2	1	2	{b,c}
EC4	3	3	1	3	{a}
EC5	2	2	2	3	{a,b,c}
EC6	1	2	2	2	{a}

Sehingga Rules yang terbentuk adalah sbb :

1. If A = 2 and B = 3 then D = 1 atau If Jenis bangunan = "Permanen" and Luas Bangunan = 60 m2 s/d 100 m2 then Tarif Naik 5 %
2. If B = 3 and C = 2 then D = 1 atau If Luas Bangunan 60 m2 s/d 100 m2 and Letak Bangunan Jalan Lingkungan then Tarif Naik 5 %

3. If B= 2 and C= 1 then D = 2 atau If Luas Bangunan 60 m2 s/d 100 m2 and Letak Bangunan Jalan Utama then Tarif Naik 10%
4. If A = 3 then D = 3 atau Jika Jenis Bangunan = Semi Permanen then Tarif Naik 5 %
5. If A = 2 ,B = 2 and C = 2 then D =3 atau If Jenis Bangunan = Permanen, Luas Bangunan 60 m2 s/d 100 m2, and Letak Bangunan = Jalan Lingkungan then Tarif Naik 5%
6. If A =1 then D = 2 atau If Jenis bangunan = "Mewah" then Tarif Naik 10%

Sedangkan *Knowledge ( Rules)* dari *ekstraksi data mining* yang terbentuk berdasarkan *equivalence class* yang terbentuk untuk metode algoritma *Mean Interval* adalah sebagai berikut :

**Tabel 3-14 Generating Rules**

Class	A	B	C	D	Reduct
EC1	2	3	2	1	{a,b}, {b}
EC2	2	3	1	1	{a,b}, {b}
EC3	2	2	1	2	{a,b,c}
EC4	2	2	2	3	{a,b,c}
EC5	3	2	1	3	{a}
EC6	1	2	2	2	{a}

Sehingga Rules yang terbentuk adalah sbb :

1. If A = 2 and B = 3 then D = 1 atau If Jenis bangunan = "Permanen" and Luas Bangunan = 60 m2 s/d 100 m2 then Tarif Naik 5 %
2. If B = 3 then D = 1 atau If Luas Bangunan 60 m2 s/d 100 m2 then Tarif Naik 5 %
3. If A = 2, B = 2 and C= 1 then D = 2 atau If Jenis Bangunan = Permanen, Luas Bangunan 40m2 s/d 60 m2 and Letak Bangunan Kelas Jalan Utama then Tarif Naik 10 %
4. If A = 2, B = 2 and C= 2 then D = 3 atau If Jenis Bangunan = Permanen, Luas Bangunan = 40m2 s/d 60 m2 and Letak Bangunan = Kelas Jalan Utama then Tarif Naik 5 %
5. If A = 3 Then D = 3 atau If Jenis Bangunan = Semi Permanen then Tarif Naik 5 %
6. If A = 1 Then D = 2 atau If Jenis Bangunan = Mewah then Tarif Naik 10 %

## 6. Kesimpulan dan Saran

### 6.1 Kesimpulan

Kesimpulan yang diperoleh selama melaksanakan penelitian ini adalah :

1. Algoritma teknik *Data Cleaning* merupakan algoritma yang mengandalkan proses yang mampu menangani data tidak lengkap data dalam pembangunan struktur data dari data tak lengkap yang menghasilkan *Knowledge* sebagai bentuk akhir keluarannya yang dapat mendukung pengambilan keputusan.
2. Dalam penelitian ini telah dilakukan perbandingan dua buah metode *Data Cleaning* yaitu algoritma *Mean Substitusi* dan algoritma *Mean Interval* yang penulis temukan sendiri.
3. Algoritma *Mean Interval* dari teknik *Data Cleaning* ternyata lebih efektif dibanding teknik data *Mean Substitusi*, karena *search space* dari algoritma ini dapat dipersempit sehingga pencarian lebih efisien
4. Struktur yang dihasilkan algoritma ini sangat dipengaruhi oleh jumlah data
1. Mencoba membandingkan dan menemukan teknik baru dari data *Cleaning*, teknik mana yang lebih efektif.
2. Mencoba melakukan pengujian antara algoritma *Mean Substitusi* dan *Mean Interval* terhadap kasus data tidak lengkap dalam *Data Mining*.
3. Mencoba melakukan pengujian bagaimana *Rough Set* dapat membantu pencarian *knowledge* dari data tidak lengkap dan menyajikannya dalam bentuk *Rules Generating*.
4. Mencoba melakukan pengujian performansi algoritma untuk kasus *estimasi* kenaikan tarif PDAM Sungai Penuh, dan karena keterbatasan waktu dan pengetahuan belum dilakukan pembandingan terutama untuk kasus-kasus lain yang lebih besar baik dari segi jumlah data maupun jumlah atribut.

## 6.2 Saran

Saran yang diharapkan dapat membantu terutama untuk pengembangan penelitian ini nantinya adalah :

## 5. Referensi

- [1] Ariasih dkk. 2011. Pengantar Data Mining. Program Studi Magister Teknik Elektro, Universitas Udayana. Denpasar
- [2] Buaton Relita. 2014. 15 Metode Menyelesaikan Data Mining, Sistem Pakar dan Sistem Pendukung Keputusan. STMIK Kaputama. Sumatera Utara
- [3] Budi Santosa, 2007 Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis. Teori dan Aplikasi . Yogyakarta, Graha Ilmu
- [4] Dahlan Abdullah, Cut Ita Erliana, Juliana "Implementasi Metode Rough Set Untuk Menentukan Data Nasabah Potensial Mendapatkan Pinjaman" Prosiding SENATKOM 2015 ISSN : 2460-4690
- [5] Dunham, Margaret H. 2003. Data Mining Introductory and Advanced Topics, New Jersey: Prentice Hall
- [6] Erhard Rahm, H. Do, Data Cleaning: Problems and Current Approaches, IEEE Data Engineering Bulletin, 2000, Vol.4 pp.3-13,
- [7] Han, J. et al. 2006 Data Mining; Concepts and Techniques, 2nd Edition, Morgan Kaufmann Publisher,
- [8] Kusriani, Emha Taufiq Luthfi. 2009 Algoritma Data Mining .Yogyakarta : Andi Offset
- [9] Larose , Daniel T, 2005, Discovering Knowledge in Data: An Introduction to Data Mining, John Wiley & Sons. Inc
- [10] Listiana Nila, Anggraini W, Achmad Mukhlason. 2012. Implementasi

Algoritma Rough Set Untuk Deteksi  
Dan Penanganan Dini Penyakit Sapi.  
Jurnal. Fakultas Teknologi Informasi.  
Institut Teknologi Sepuluh Nopember  
Surabaya

- [11] Nofriansyah Dicky. 2014. Konsep Data Mining Vs Sistem Pendukung Keputusan. Penerbit Deepublish. Yogyakarta.
- [12] Sarjon Defit, Mohd. Noor Md. Sap ; 2003 Intelligent Mining Multi Dimensional Association Rules From Large Inconsistent Database ;Journal Teknologi Maklumat, Vol 15 No 1, Juni 2003
- [13] Turban, Efraim, dkk. 2005. Decision Support Systems and Intelligent Systems (Sistem Pendukung Keputusan dan Sistem Cerdas) Edisi 7. Yogyakarta : Andi Offset.
- [14] Witten, Ian H, And Eibe Frank, 2005 Data Mining : Practical Machine Learning Tools And Techniques, 2nd Edition, Morgan Kaufmann Publisher.
- [15] ....., Air Minum dan Penyehatan Lingkungan (AMPL) kabupaten Kerinci 2004
- [16] <http://www.ampl.or.id/digilib/read/penetapan-tarif-air-minum-perusahaan-daerah-air-minum-tirta-sakti-kabupaten-kerinci/47550> diakses 20 Juni 2016