PENERAPAN ROUGH SET DAN FUZZY ROUGH SET UNTUK KLASIFIKASI DATA TIDAK LENGKAP

Oleh:

Winda Aprianti

NRP. 1213 201 029

Dosen Pembimbing :

Dr. Imam Mukhlash, S.Si, M.T

PROGRAM PASCA SARJANA

JURUSAN MATEMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

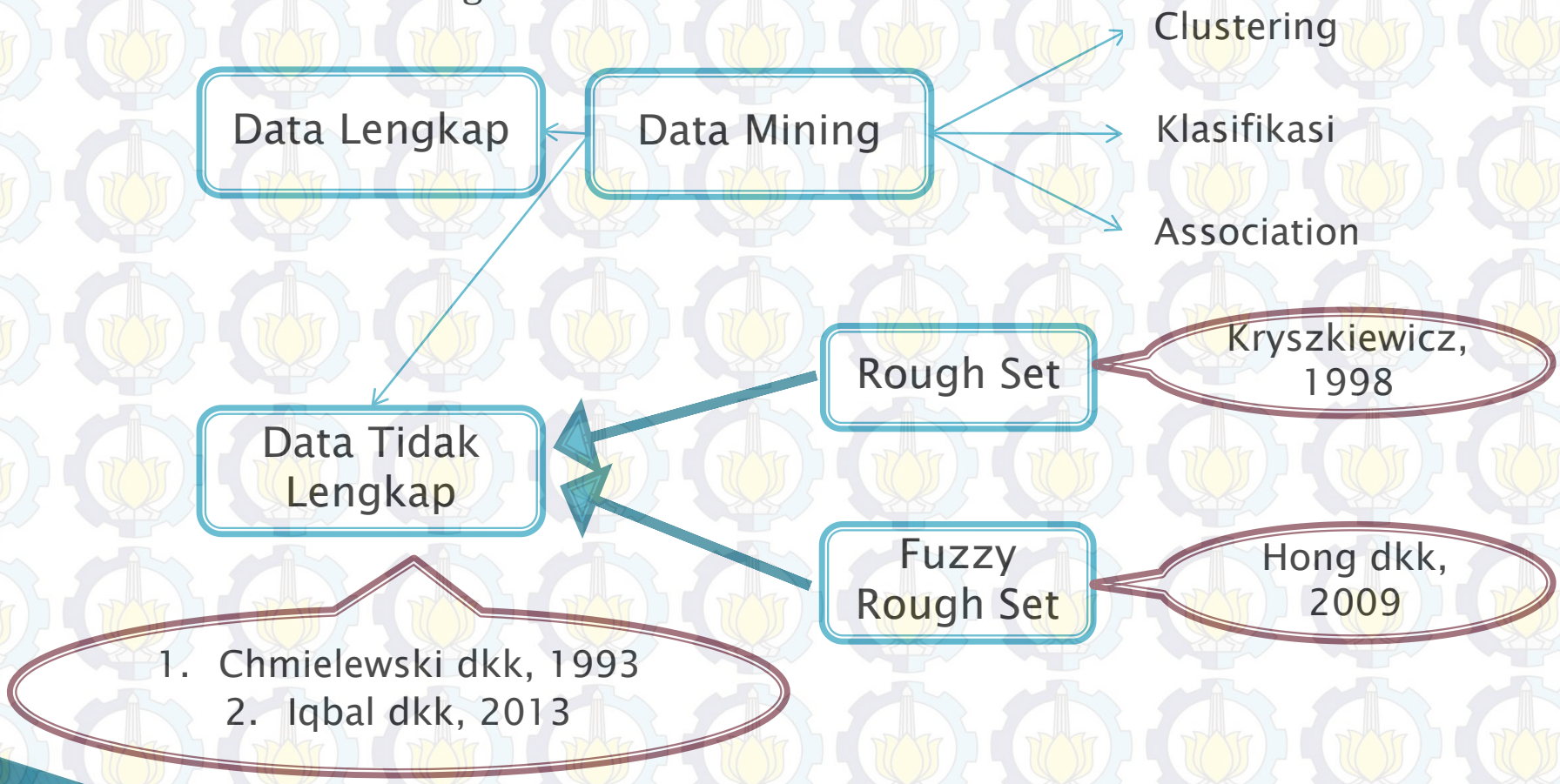
INSTITUT TEKNOLOGI SEPULUH NOPEMBER

SURABAYA

2015

PENDAHULUAN

Latar belakang



PENDAHULUAN

Latar belakang

Data Cuaca Tidak Lengkap

- Rough Set
- Fuzzy Rough Set

Penggunaan Rough Set dan Fuzzy Rough Set pada Klasifikasi Data Tidak Lengkap

PENDAHULUAN

Penelitian yang Relevan

	Research
Francis and Shen	Menerapkan *rough set* pada masalah prediksi ekonomi dan keuangan berdasarkan keauratan identifaksi pola di data histori, data histori berusi multi atribut. Hasil penelitian menunjukkan bahwa *rough set* berlaku untuk masalah yang berhubungan dengan prediksi ekonomi dan keuangan.
Maharani	Hasil penelitian menunjukkan bahwa algoritma RANFIS yang terdiri dari JST, sistem *fuzzy,* dan *rough set* merupakan sistem yang mampu memprediksi nilai saham. Rough set memiliki kemampuan dalam menurunkan nilai error.
Xiao-feng and Song-song	Penggunaan *fuzzy rough set* dapat membuat hasil prediksi dari nilai saham lebih efektif.

PENDAHULUAN

Penelitian yang Relevan

	Research
Sadiq, Dualmi, and Shaker	Hasil penelitian menunjukkan bahwa pendekatan hybrid antara rough set dan swarm intelligent lebih baik dibandingkan algoritma ID3 untuk mereduksi banyaknya rules yang dihasilkan tanpa mempengaruhi akurasi dari perkiraan nilai null , terutama ketika banyakanya nilai null ditambah.
Kryszkiewizc	Meningkatkan rough set klasik untuk mengatasi data tidak lengkap dengan mendefinisikan relasi similarity.
Hong dkk	Memperkenalkan fuzzy rough set untuk memperoleh rules dari data tidak lengkap. Hong dkk mendefinisikan fuzzy incomplete equivalence class, fuzzy incomplete lower dan upper approximation.

PENDAHULUAN

Rumusan Masalah

1. Bagaimana memperoleh *rules* dari data tidak lengkap menggunakan *Rough Set*?
2. Bagaimana memperoleh *rules* dari data tidak lengkap menggunakan *Fuzzy Rough Set*?
3. Bagaimana perbandingan *rules* yang diperoleh dari kedua algoritma di atas?

Batasan Masalah

1. Data yang digunakan merupakan data cuaca di Stasiun Perak, Surabaya.
2. Atribut dari data yang digunakan adalah temperatur, kelembaban, tekanan, kecepatan angin, dan curah hujan.

PENDAHULUAN

Tujuan Penelitian

1. Membuat *rules* dari data tidak lengkap menggunakan *Rough Set*.
2. Membuat *rules* dari data tidak lengkap menggunakan *Fuzzy Rough Set*.
3. Membandingkan *rules* yang diperoleh dari kedua algoritma di atas.

Manfaat Penelitian

Manfaat penelitian yang diperoleh dari penelitian ini adalah untuk menambah wawasan keilmuan mengenai penggunaan *rough set* dan *fuzzy rough set* pada dataset cuaca tidak lengkap. Algoritma *rough set* dan *fuzzy rough set* dapat digunakan sebagai penanganan *missing value* pada dataset tidak lengkap. *Rules* yang diperoleh dari algoritma *rough set* dan *fuzzy rough set* dapat digunakan untuk membantu prediksi curah hujan pada waktu mendatang.

DASAR TEORI

Data Mining

Data mining adalah proses untuk menemukan pola yang menarik dan pengetahuan dari data dalam jumlah besar.

Task Data Mining

- Clustering
- Association Rules
- Klasifikasi
- Regression

DASAR TEORI

Klasifikasi

Klasifikasi adalah proses menemukan model atau fungsi yang menggambarkan dan membedakan kelas data atau konsep. Model ini digunakan untuk memprediksi label kelas untuk objek pada data uji (Han dkk, 2012).

Model klasifikasi dapat direpresentasikan dalam berbagai bentuk, seperti aturan klasifikasi (misalnya, aturan *If-Then*), pohon keputusan, rumus matematika, dan jaringan syaraf.

DASAR TEORI

Dataset Tidak Lengkap

Jika setidaknya satu objek di dataset mempunyai *missing value* maka diklasifikasikan sebagai dataset tidak lengkap (Hong dkk, 2009).

Beberapa pendekatan mengubah dataset tidak lengkap dapat diubah menjadi dataset lengkap dengan, seperti yang dikemukakan Grzymala-Busse (2004) dan Jiawei Han dkk (2012) sebagai berikut:

- mengganti nilai atribut yang hilang dengan nilai atribut yang paling umum (paling sering terjadi),
- untuk atribut numerik, nilai atribut yang hilang diganti dengan nilai rata-rata atribut,
- menentukan semua kemungkinan nilai atribut.
- mengabaikan kasus dengan nilai atribut hilang. mempertimbangkan nilai atribut hilang sebagai nilai khusus.
- mengganti nilai atribut yang hilang dengan rata-rata atau median dari atribut untuk semua objek yang memiliki kelas keputusan yang sama,
- mengganti nilai atribut yang hilang dengan nilai yang mungkin menggunakan regresi, alat berbasis inferensi menggunakan Bayesian, atau induksi pohon keputusan.

DASAR TEORI

Rough Set

Information System and Decision Table

Information system adalah $I = (U, A)$, dimana U adalah himpunan tidak kosong dari objek berhingga (semesta pembicaraan) dan A adalah himpunan berhingga yang tidak kosong dari atribut sehingga

$$a : U \rightarrow V_a$$

untuk setiap $a \in \mathbb{A}$. V_a adalah himpunan nilai dari atribut a.

D*ecision system*, $A = \{C \cup D\}$, dimana C adalah himpunan input fitur dan D adalah himpunan dari kelas.

DASAR TEORI

Indiscernibility Relation

Misal $v_j^{(i)}$ adalah nilai dari atribut A_j untuk objek ke-i $Obj^{(i)}$. $Obj^{(i)}$ dan $Obj^{(k)}$ dikatakan memiliki *indiscernibility relation* (atau relasi ekivalensi) pada atribut A_j, jika $Obj^{(i)}$ dan $Obj^{(k)}$ memiliki nilai atribut A_j yang sama $\left(v_j^{(i)} = v_j^{(k)}\right)$

Lower Approximation :

himpunan semua objek yang pasti diklasifikasikan sebagai suatu subset

$$\underline{B}X = \{x | x \in U, B(x) \subseteq X\}$$

Upper Approximation :

himpunan semua objek yang mungkin diklasifikasikan sebagai subset

$$\overline{B}X = \{x | x \in U \; dan \; B(x) \cap X \neq \emptyset\}$$

DASAR TEORI

Incomplete Information System

Information system yang memiliki setidaknya satu nilai hilang untuk atribut dari objek disebut *incomplete information system*. Nilai atribut yang hilang dinotasikan dengan simbol $*$.

Sedangkan *incomplete decision table* adalah $I = (U, \{C \cup D\})$, dimana $d \in D, d \notin C$, dan $* \notin V_d$ (Kryszkiewicz,1998)

DASAR TEORI

Similarity Relation

Kryszkiewicz (1998) mengusulkan pendekatan rough *set* untuk langsung mempelajari *rules* dari dataset yang tidak lengkap, dengan cara mendefinisikan *similarity relation* sebagai berikut.

$$SIM(A) = \{(x, y) \in U \times U | a \in A, a(x) = a(y)\ atau\ a(x) =* atau\ a(y) =*\}$$

DASAR TEORI

Algoritma Klasifikasi Rough Set

DASAR TEORI

Fuzzy Rough Set

Fuzzy Incomplete Equivalence Class

Jika objek $Obj^{(i)}$ memiliki nilai *uncertain(u)* untuk atribut A_j, maka $(Obj^{(i)}, u)$ dimasukkan ke dalam setiap kelas ekivalensi *fuzzy* dari atribut A_j.

Jika objek $Obj^{(i)}$ memiliki nilai keanggotaan fuzzy *certain* $f_{j_k}^{(i)}$ untuk atribut A_j, masukkan $(Obj^{(i)}, c)$ ke dalam *fuzzy incomplete equivalence class* dari $A_j = R_{j_k}$

Derajat keanggotaan $\mu_{A_{j_k}} = \min_i f_{j_k}^{(i)}, f_{j_k}^{(i)} \neq 0.$

DASAR TEORI

Fuzzy Incomplete Lower dan Upper Approximation

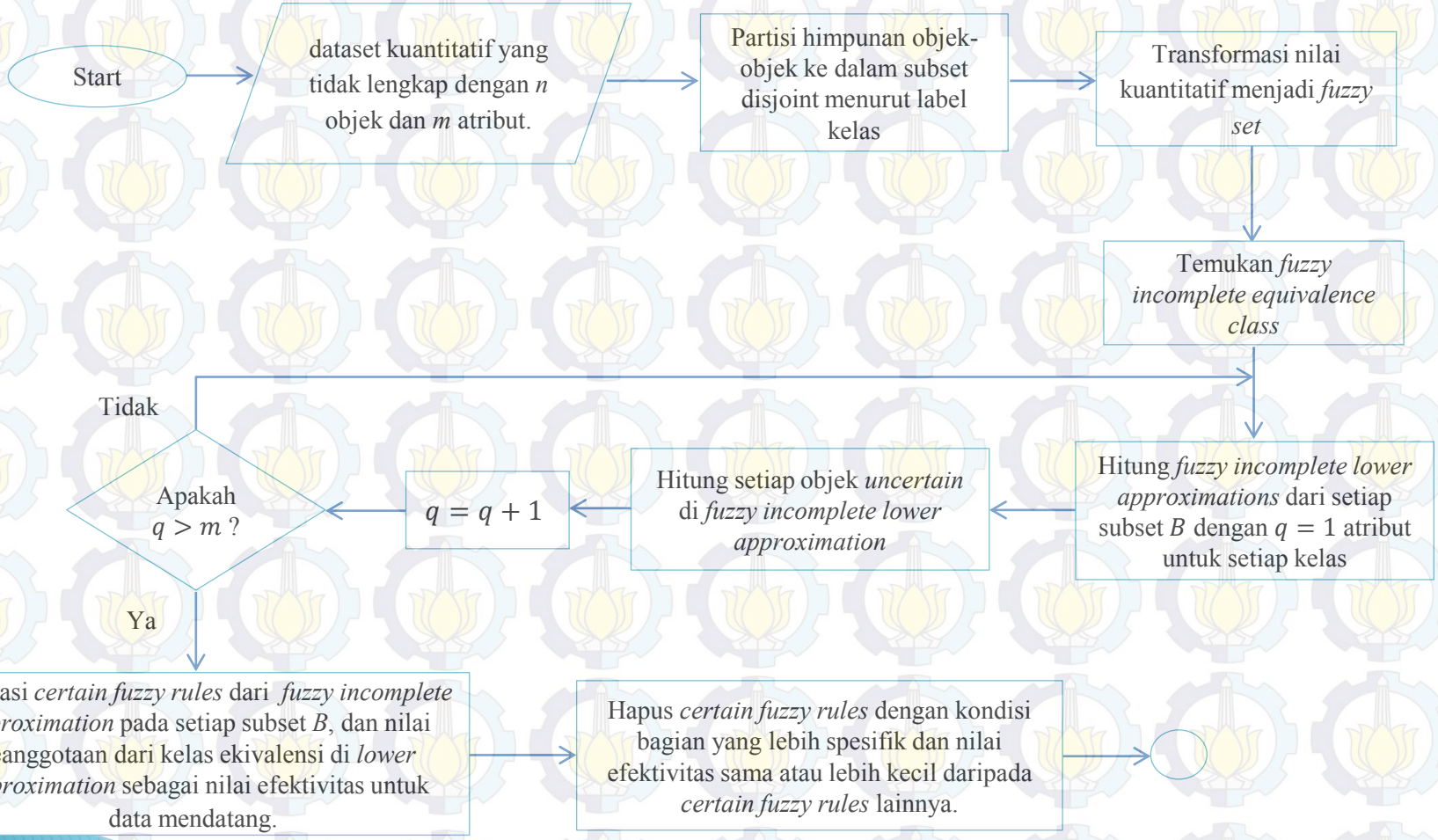
Fuzzy incomplete lower dan upper approximation didefinisikan sebagai berikut.

- $\underline{B}X_l = \{(B_k(Obj^{(i)}), \mu_{B_k}(Obj^{(i)}))|1 \leq i \leq n, Obj^{(i)} \in X_l, B_k^c(Obj^{(i)}) \subseteq X_l, 1 \leq k \leq |B(Obj^{(i)})|\}$

- $\overline{B}X_l = \{(B_k(Obj^{(i)}), \mu_{B_k}(Obj^{(i)}))|1 \leq i \leq n, B_k^c(Obj^{(i)}) \cap X_l \neq \emptyset, B_k^c(Obj^{(i)}) \not\subseteq X_l, 1 \leq k \leq |B(Obj^{(i)})|\}$

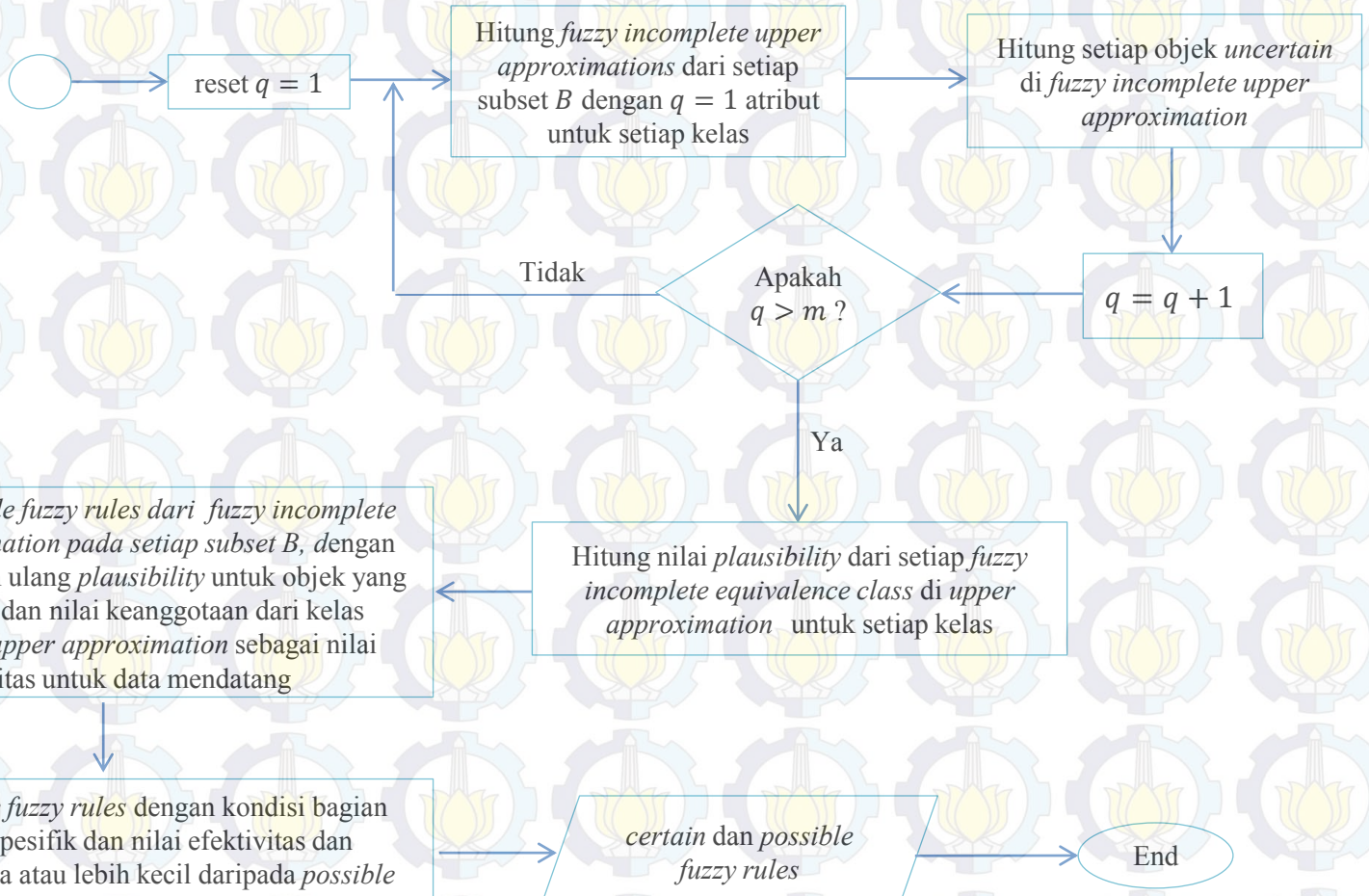
DASAR TEORI

Algoritma Klasifikasi Fuzzy Rough Set



DASAR TEORI

Algoritma Klasifikasi Fuzzy Rough Set



DASAR TEORI

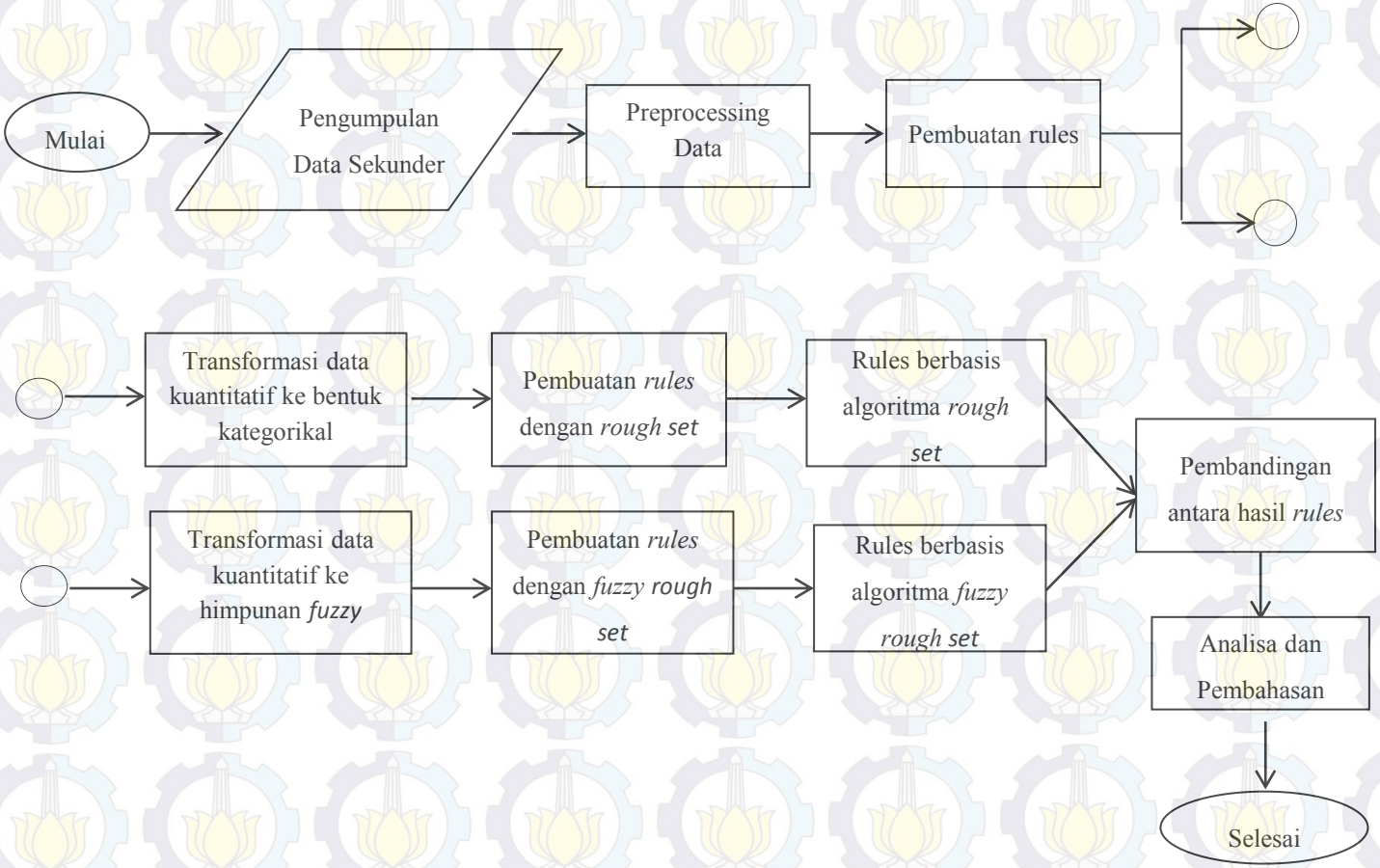
Cuaca

Menurut kamus besar bahasa Indonesia, cuaca adalah keadaan udara pada satu tempat tertentu dengan jangka waktu terbatas. Keadaan cuaca senantiasa berubah dari waktu ke waktu (Pusat Bahasa Departemen Pendidikan Nasional RI, 2014)

Keadaan cuaca dipengaruhi oleh faktor-faktor alamiah berupa suhu atau temperatur udara, tekanan udara, angin, kelembaban udara, dan curah hujan

METODE PENELITIAN

Diagram alir penelitian



HASIL DAN PEMBAHASAN

Pengumpulan Dataset

Dataset yang digunakan merupakan data sekunder meteorologi di stasiun Perak, Surabaya pada tahun 2005-2009 dari Badan Meteorologi, Klimatologi, dan Geogologi Indonesia.

Dataset terdiri dari 5 atribut, yaitu temperatur rata-rata, kelembaban, tekanan udara, kecepatan angin, dan curah hujan, dimana curah hujan merupakan atribut keputusan.

Dataset terdiri dari 1826 objek.

HASIL DAN PEMBAHASAN

Klasifikasi Algoritma Rough Set

Kategorisasi Atribut Curah Hujan

Nilai numerik dari atribut curah hujan diubah dalam bentuk nilai kategorikal berdasarkan 5 kategori berikut.

$$Curah\ Hujan\ (CH) = \begin{cases} Sangat\ Ringan\ (SR) & , CH < 5 \\ Ringan\ (R) & , 5 \leq CH < 20 \\ Normal\ (N) & , 20 \leq CH < 50 \\ Lebat\ (L) & , 50 \leq CH < 100 \\ Sangat\ Lebat\ (SL) & , CH \geq 100 \end{cases}$$

HASIL DAN PEMBAHASAN

Kategorisasi Atribut

$$Temperatur\ (A_1) = \begin{cases} Dingin\ (D) & , A_1 \leq 26.5 \\ Normal\ (N) & , 26.5 < A_1 < 29 \\ Panas\ (P) & , A_1 \geq 29 \end{cases}$$

$$Kelembaban\ (A_2) = \begin{cases} Kering\ (K) & , A_2 \leq 68 \\ Lembab\ (L) & , 68 < A_2 < 78 \\ Basah\ (B) & , A_2 \geq 78 \end{cases}$$

$$Tekanan\ udara\ (A_3) = \begin{cases} Rendah\ (R) & , A_3 \leq 1008 \\ Normal\ (N) & , 1008 < A_3 < 1013 \\ Tinggi\ (T) & , A_3 \geq 1013 \end{cases}$$

$$Kecepatan\ angin\ (A_4) = \begin{cases} Lambat\ (L) & , A_4 \leq 4 \\ Normal\ (N) & , 4 < A_4 < 8 \\ Kencang\ (K) & , A_4 \geq 8 \end{cases}$$

HASIL DAN PEMBAHASAN

Proses Klasifikasi dengan Algoritma Rough Set

Tabel 4.1 Incomplete Decision Table

Objek	A_1	A_2	A_3	A_4	CH
1	27.7	92.1	1006.9	*	1
2	26.3	94.1	*	4.8	70.1
3	*	93.9	1006.4	7.1	59.9
4	27.5	94.3	1007.3	3.2	37.1
5	24	96.9	1008	5.5	8.9
6	26.1	*	1009.1	1.6	17
7	*	95.1	1012.7	6.4	33
8	27.9	67.2	1011.7	10.3	0
9	29	74.5	1012	12.6	0
10	31.9	82.4	*	6.8	0

Objek	A_1	A_2	A_3	A_4	CH
1	Normal	Basah	Rendah	*	SR
2	Sejuk	Basah	*	Normal	L
3	*	Basah	Rendah	Normal	L
4	Normal	Basah	Rendah	Lambat	N
5	Sejuk	Basah	Rendah	Normal	R
6	Sejuk	*	Sedang	Lambat	R
7	*	Basah	Sedang	Normal	N
8	Normal	Kering	Sedang	Kencang	SR
9	Panas	Lembab	Sedang	Kencang	SR
10	Panas	Basah	*	Normal	SR

HASIL DAN PEMBAHASAN

Proses Klasifikasi dengan Algoritma Rough Set

Tabel Matriks *discernibility*.

Objek	1	2	3	4	5	6	7	8	9	10
1	-	A_1	-	-	A_1	A_1A_3	A_3	-	-	-
2	A_1	-	-	A_1A_4	-	A_4	-	$A_1A_2A_4$	$A_1A_2A_4$	A_1
3	-	-	-	A_4	-	A_3A_4	A_3	$A_2A_3A_4$	$A_2A_3A_4$	-
4	-	A_1A_4	A_4	-	A_1A_4	A_1A_3	-	$A_2A_3A_4$	$A_1A_2A_3A_4$	A_1A_4
5	A_1	-	-	A_1A_4	-	-	A_3	$A_1A_2A_3A_4$	$A_1A_2A_3A_4$	A_1
6	A_1A_3	A_4	A_3A_4	A_1A_3	-	-	A_4	A_1A_4	A_1A_4	A_1A_4
7	A_3	-	A_3	-	A_3	A_4	-	A_2A_4	A_2A_4	-
8	-	$A_1A_2A_4$	$A_2A_3A_4$	$A_2A_3A_4$	$A_1A_2A_3A_4$	A_1A_4	A_2A_4	-	-	-
9	-	$A_1A_2A_4$	$A_2A_3A_4$	$A_1A_2A_3A_4$	$A_1A_2A_3A_4$	A_1A_4	A_2A_4	-	-	-
10	-	A_1	-	A_1A_4	A_1	A_1A_4	-	-	-	-

Misal fungsi *discernibility untuk objek 1*:

$$A_1 \wedge A_1 \wedge (A_1 \vee A_3) \wedge A_3 = (A_1 \wedge A_3)$$

Dari fungsi *discernibility* untuk setiap objek pada diperoleh *incomplete reduced decision table* yang berisikan 4 atribut, yaitu A_1, A_3, A_4, dan CH.

HASIL DAN PEMBAHASAN

Proses Klasifikasi dengan Algoritma Rough Set

Matriks *discernibility* dari *incomplete reduced decision table*

Objek	1	2	3	4	5	6	7	8	9	10
1	-	A_1	-	-	A_1	A_1A_2	A_3	-	-	-
2	A_1	-	-	A_1A_4	-	A_4	-	A_1A_4	A_1A_4	A_1
3	-	-	-	A_4	-	A_2A_4	A_3	A_2A_4	A_2A_4	-
4	-	A_1A_4	A_4	-	A_1A_4	A_1A_2	-	A_2A_4	$A_1A_2A_4$	A_1A_4
5	A_1	-	-	A_1A_4	-	-	A_3	$A_1A_2A_4$	$A_1A_2A_4$	A_1
6	A_1A_2	A_4	A_2A_4	A_1A_2	-	-	A_4	A_1A_4	A_1A_4	A_1A_4
7	A_3	-	A_3	-	A_3	A_4	-	A_4	A_4	-
8	-	A_1A_4	A_2A_4	A_2A_4	$A_1A_2A_4$	A_1A_4	A_4	-	-	-
9	-	A_1A_4	A_2A_4	$A_1A_2A_4$	$A_1A_2A_4$	A_1A_4	A_4	-	-	-
10	-	A_1	-	A_1A_4	A_1	A_1A_4	-	-	-	-

Hasil *reduct* pada matriks discernibility di atas menghasilkan atribut A_1, A_3, dan A_4

HASIL DAN PEMBAHASAN

Proses Klasifikasi dengan Algoritma Rough Set

Tabel *Rules* dari *Incomplete Reduce Decision Table*.

Objek	Fungsi *Discernibility*	*Rules*	p
1	$(A_1 \wedge A_3)$	Jika (A_1 = Normal) dan (A_3 = Rendah) maka CH = SR	0.5
2	$(A_1 \wedge A_4)$	Jika (A_1 = Sejuk) dan (A_4 = Normal) maka CH = L	0.5
3	$(A_3 \wedge A_4)$	Jika (A_3 = Rendah) dan (A_4 = Normal) maka CH = L	0.5
4	$A_4 \wedge (A_1 \vee A_3)$	Jika (A_1 = Normal) dan (A_4 = Lambat) maka CH = N	1
		Jika (A_3 = Rendah) dan (A_4 = Lambat) maka CH = N	1
5	$(A_1 \wedge A_3)$	Jika (A_1 = Sejuk) dan (A_3 = Rendah) maka CH = R	1
6	$A_4 \wedge (A_1 \vee A_3)$	Jika (A_1 = Sejuk) dan (A_4 = Lambat) maka CH = R	1
		Jika (A_3 = Sedang) dan (A_4 = Lambat) maka CH = R	1
7	$(A_3 \wedge A_4)$	Jika (A_3 = Sedang) dan (A_4 = Normal) maka CH = N	1
8	A_4	Jika (A_4 = Kencang) maka CH = SR	1
9	A_4	Jika (A_4 = Kencang) maka CH = SR	1
10	A_1	Jika (A_1 = Panas) maka CH = SR	

Untuk setiap *rules*, dicari nilai *plausibility* dengan rumus berikut.

$$P\left(B_k\left(Obj^{(i)}\right)\right) = \frac{\left|B_k\left(Obj^{(i)}\right) \cap X_l\right|}{\left|B_k\left(Obj^{(i)}\right)\right|}$$

HASIL DAN PEMBAHASAN

Proses Klasifikasi Algoritma Fuzzy Rough Set

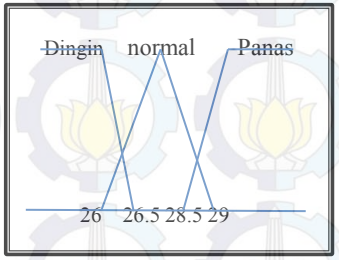
Kategorisasi dan Fuzzifikasi Data

Mengubah atribut curah hujan ke dalam nilai kategorikal

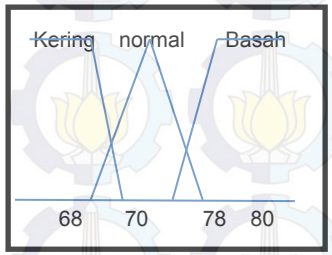
Mempartisi himpunan objek-objek berdasarkan 5 kelas keputusan

X_{SR}= {1,8,9,10}, X_R = {5,6}, X_N = {4,7}, X_L = {2,3}, dan X_{SL} = { }.

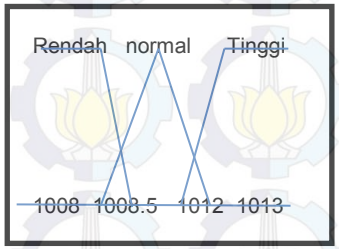
Fuzzifikasi untuk atribut selain atribut curah hujan dengan menggunakan fungsi keanggotaan berikut.



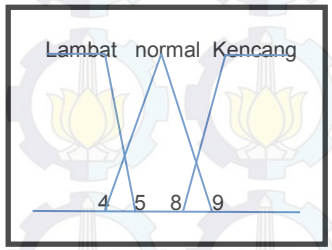
(a)



(b)



(c)



(d)

HASIL DAN PEMBAHASAN

Proses Klasifikasi dengan Algoritma Fuzzy Rough Set

Incompleted decision table dalam fuzzy set:

Objek	A_1	A_2	A_3	A_4	CH
1	$\frac{0.87}{normal}$	$\frac{1}{basah}$	$\frac{1}{rendah}$	*	SR
2	$\frac{0.4}{sejuk}+\frac{0.2}{normal}$	$\frac{1}{basah}$	*	$\frac{0.2}{lambat}+\frac{0.32}{AK}$	L
3	*	$\frac{1}{basah}$	$\frac{1}{rendah}$	$\frac{0.76}{AK}$	L
4	$\frac{1}{normal}$	$\frac{1}{basah}$	$\frac{1}{rendah}$	$\frac{1}{lambat}$	N
5	$\frac{1}{sejuk}$	$\frac{1}{basah}$	$\frac{1}{rendah}$	$\frac{0.6}{AK}$	R
6	$\frac{0.8}{sejuk}+\frac{0.07}{normal}$	*	$\frac{0.44}{sedang}$	$\frac{1}{lambat}$	R
7	*	$\frac{1}{basah}$	$\frac{0.12}{sedang}+\frac{0.2}{tinggi}$	$\frac{0.96}{AK}$	N
8	$\frac{0.73}{normal}$	$\frac{1}{kering}$	$\frac{0.52}{sedang}$	$\frac{1}{kencang}$	SR
9	$\frac{1}{panas}$	$\frac{0.92}{lembab}$	$\frac{0.4}{sedang}$	$\frac{1}{kencang}$	SR
10	$\frac{1}{panas}$	$\frac{1}{basah}$	*	$\frac{0.53}{AK}$	SR

HASIL DAN PEMBAHASAN

Proses Klasifikasi dengan Algoritma Fuzzy Rough Set

Contoh :

Fuzzy incomplete equivalence class untuk atribut A_4

$$U/A_4 = \{\{(Obj^{(1)},u)(Obj^{(2)},c)(Obj^{(4)},c)(Obj^{(6)},c),0.2\},\{(Obj^{(1)},u)(Obj^{(2)},c)(Obj^{(3)},c)(Obj^{(5)},c)(Obj^{(7)},c)(Obj^{(10)},c),0.32\},\{(Obj^{(1)},u)(Obj^{(8)},c)(Obj^{(9)},c),1\}\}$$

Fuzzy incomplete lower approximation untuk atribut A_4

$$\underline{A_4}(X_{VSR}) = \{(Obj^{(1)},u)(Obj^{(8)},c)(Obj^{(9)},c),1\}$$

$$\underline{A_4}(X_{SR}) = \underline{A_4}(X_R) = \underline{A_4}(X_{HR}) = \emptyset$$

HASIL DAN PEMBAHASAN

Proses Klasifikasi dengan Algoritma Fuzzy Rough Set

Pada $\underline{A_4}(X_{SR})$, karena $(1, u)$ hanya ada di satu *fuzzy incomplete equivalence class* dari $A_4 = kencang$ maka nilai dari objek (1) dapat diperkirakan sebagai berikut

$$\frac{10.3(1) + 12.6(1)}{1 + 1} = 11.45 \rightarrow \frac{1}{kencang}$$

Dengan mengganti nilai $(\mathrm{Obj}^{(1)}, \mathrm{u})$ diperoleh

$$\mathrm{U}/A_4 = \{\{(\mathrm{Obj}^{(2)}, \mathrm{c})(\mathrm{Obj}^{(4)}, \mathrm{c})(\mathrm{Obj}^{(6)}, \mathrm{c}), 0.2\}, \{(\mathrm{Obj}^{(2)}, \mathrm{c})(\mathrm{Obj}^{(3)}, \mathrm{c})(\mathrm{Obj}^{(5)}, \mathrm{c})(\mathrm{Obj}^{(7)}, \mathrm{c})(\mathrm{Obj}^{(10)}, \mathrm{c}), 0.32\}, \{(\mathrm{Obj}^{(1)}, \mathrm{c})(\mathrm{Obj}^{(8)}, \mathrm{c})(\mathrm{Obj}^{(9)}, \mathrm{c}), 1\}\}$$

$$\underline{A_4}(\mathrm{X_{SR}}) = \{(\mathrm{Obj}^{(1)}, \mathrm{c})(\mathrm{Obj}^{(8)}, \mathrm{c})(\mathrm{Obj}^{(9)}, \mathrm{c}), 1\}$$

$$\underline{A_4}(\mathrm{X_R}) = \underline{A_4}(\mathrm{X_N}) = \underline{A_4}(X_L) = \emptyset$$

HASIL DAN PEMBAHASAN

Proses Klasifikasi dengan Algoritma Fuzzy Rough Set

Modifikasi dari *Incompleted decision table* dalam *fuzzy set*:

A_1	A_2	A_3	A_4	CH
$\frac{0.87}{normal}$	$\frac{1}{basah}$	$\frac{1}{rendah}$	$\frac{1}{kencang}$	SR
$\frac{0.4}{sejuk}+\frac{0.2}{normal}$	$\frac{1}{basah}$	$\frac{0.12}{Sedang}+\frac{0.2}{tinggi}$	$\frac{0.2}{lambat}+\frac{0.32}{AK}$	L
$\frac{1}{panas}$	$\frac{1}{basah}$	$\frac{1}{rendah}$	$\frac{0.76}{AK}$	L
$\frac{1}{normal}$	$\frac{1}{basah}$	$\frac{1}{rendah}$	$\frac{1}{lambat}$	N
$\frac{1}{sejuk}$	$\frac{1}{basah}$	$\frac{1}{rendah}$	$\frac{0.6}{AK}$	R
$\frac{0.8}{sejuk}+\frac{0.07}{normal}$	$\frac{1}{kering}$	$\frac{0.44}{sedang}$	$\frac{1}{lambat}$	R
$\frac{1}{panas}$	$\frac{1}{basah}$	$\frac{0.12}{Sedang}+\frac{0.2}{tinggi}$	$\frac{0.96}{AK}$	N
$\frac{0.73}{normal}$	$\frac{1}{kering}$	$\frac{0.52}{sedang}$	$\frac{1}{kencang}$	SR
$\frac{1}{panas}$	$\frac{0.92}{lembab}$	$\frac{0.4}{sedang}$	$\frac{1}{kencang}$	SR
$\frac{1}{panas}$	$\frac{1}{basah}$	$\frac{0.12}{Sedang}+\frac{0.2}{tinggi}$	$\frac{0.53}{AK}$	SR

HASIL DAN PEMBAHASAN

Proses Klasifikasi dengan Algoritma Fuzzy Rough Set

Berdasarkan *fuzzy incomplete lower approximation* untuk 1 atribut, 2 atribut, 3 atribut dan 4 atribut diperoleh *certain rules* berikut.

Tabel *Certain Rules* dari Contoh 10 Data

No.	*Certain Rules*
1.	Jika A_2 = Normal maka CH = SR, dengan Fe = 0.92
2.	Jika A_4 = Kencang maka CH = SR, dengan Fe = 1
3.	Jika A_1 = Dingin dan A_2 = Kering maka CH = R, dengan Fe = 0.8
⋮	⋮
14.	Jika A_1 = Dingin, A_2 = Basah dan A_4 = Normal maka CH = L, dengan Fe = 0.2
15.	Jika A_1 = Dingin, A_3 = Normal dan A_4 = Normal maka CH = L, dengan Fe = 0.12
16.	Jika A_2 = Basah, A_3 = Normal dan A_4 = Lambat maka CH = L, dengan Fe = 0.12

HASIL DAN PEMBAHASAN

Proses Klasifikasi dengan Algoritma Fuzzy Rough Set

- Menemukan *fuzzy incomplete upper approximation*

Contoh :

Fuzzy incomplete upper approximation untuk atribut A_1

$$\overline{A_1}(X_{SR}) = \{\{(1,c)(2,c)(4,c)(6,c)(8,c),0.07\},\{(3,c)(7,c)(9,c)(10,c),1\}\}$$
$$\overline{A_1}(X_R) = \{\{(1,c)(2,c)(4,c)(6,c)(8,c),0.07\},\{(2,c)(5,c)(6,c),0.4\}\}$$
$$\overline{A_1}(X_N) = \{\{(1,c)(2,c)(4,c)(6,c)(8,c),0.07\},\{(3,c)(7,c)(9,c)(10,c),1\}\}$$
$$\overline{A_1}(X_L) = \{\{(1,c)(2,c)(4,c)(6,c)(8,c),0.07\},\{(2,c)(5,c)(6,c),0.4\},$$
$$\{(3,c)(7,c)(9,c)(10,c),1\}\}$$

HASIL DAN PEMBAHASAN

Proses Klasifikasi dengan Algoritma Fuzzy Rough Set

Perhitungan nilai *plausibility* pada *fuzzy incomplete upper approximation* menggunakan rumus berikut.

$$P\left(B_k^c\left(Obj^{(i)}\right)\right) = \frac{\sum_{Obj^{(r)} \in B_k^c\left(Obj^{(i)}\right) \,\&\, Obj^{(r)} \in X_l} f_{jk}^{(r)}}{\sum_{Obj^{(r)} \in B_k^c\left(Obj^{(i)}\right)} f_{jk}^{(r)}} \quad (4.2)$$

Misal perhitungan nilai *plausibility* untuk $A_{1_{SR}}(1,2,4,6,8)$:

$$P\left(A_{1_{SR}}(1,2,4,6,8)\right) = \frac{\sum_{1,8} f_{jk}^{(r)}}{\sum_{1,2,4,6,8} f_{jk}^{(r)}} = \frac{0.87 + 0.73}{0.87 + 0.2 + 1 + 0.07 + 0.73}$$
$$= 0.56$$

HASIL DAN PEMBAHASAN

Proses Klasifikasi dengan Algoritma Fuzzy Rough Set

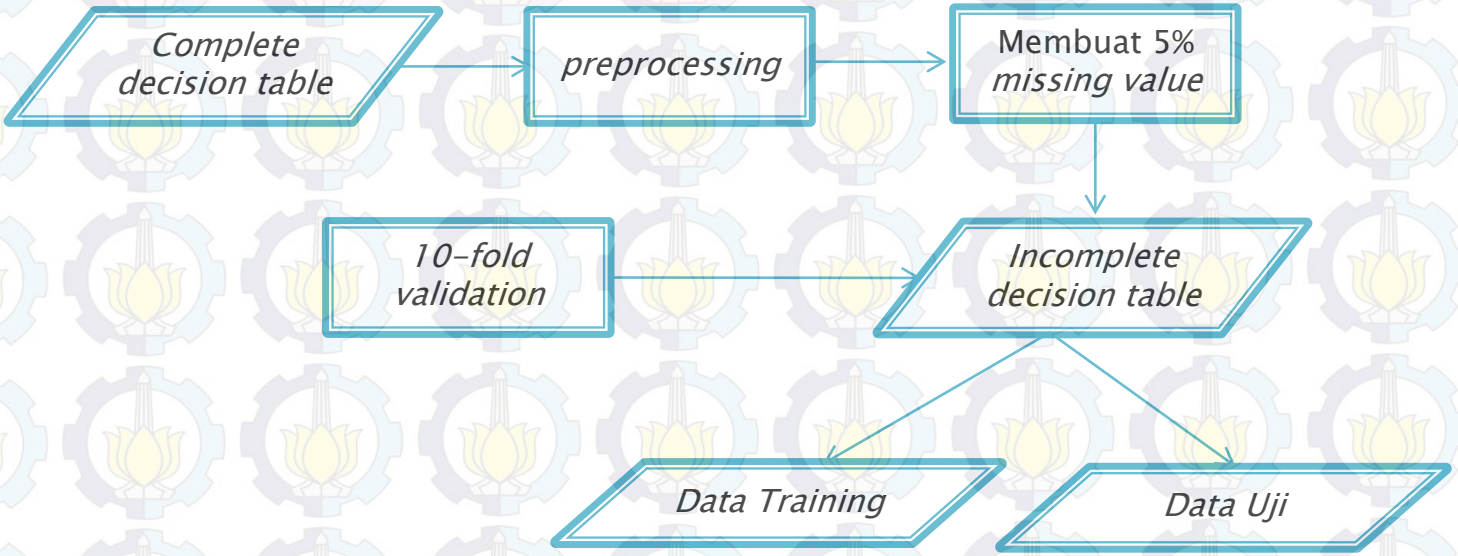
Berdasarkan *fuzzy incomplete upper approximation* untuk 1 atribut, 2 atribut, 3 atribut dan 4 atribut diperoleh *possible rules* berikut.

Certain Rules dari Contoh 10 Data

No.	*Possible Rules*
1.	Jika A_1 = Normal maka CH = SR, dengan Fe = 0.07, p = 0.56
2.	Jika A_1 = Normal maka CH = R, dengan Fe = 0.07, p = 0.02
3.	Jika A_1 = Normal maka CH = N, dengan Fe = 0.07, p = 0.35
⋮	⋮
62.	Jika A_1 = Panas, A_2 = Basah, dan A_3 = Normal maka CH = N, dengan Fe = 0.12, p = 0.5
63.	Jika A_1 = Panas, A_3 = Normal, dan A_4 = Normal maka CH = N, dengan Fe = 0.12, p = 0.5
64.	Jika A_1 = Normal, A_3 = Normal, A_4 = Lambat maka CH = L, dengan Fe = 0.07, p = 0.64

HASIL DAN PEMBAHASAN

Pengujian Hasil



HASIL DAN PEMBAHASAN

Pengujian Hasil

Pengimplementasi algoritma *rough set* dengan *Matlab* menghasilkan *rules* yang rata-rata berjumlah 120.

Rumus untuk pemilihan *rules* (Grzymala-Busse dan Hu, 2001):

$$\sum_{\text{rules yang sesuai}} Matching_{factor} * Plausibility$$

HASIL DAN PEMBAHASAN

Pengujian Hasil

Pengimplementasian algoritma *fuzzy rough set* dengan *Matlab* menghasilkan *rules* yang rata-rata berjumlah 162.

Rumus untuk pemilihan *rules* (Grzymala-Busse dan Hu, 2001):

$$\sum_{\text{rules yang sesuai}} Matching_{factor} * FutureEffectiveness * Plausibility$$

HASIL DAN PEMBAHASAN

Analisa Hasil

Tabel Perbandingan Performansi *Rules* Berbasis Algoritma *Rough Set* dan *Fuzzy Rough Set* pada Dataset dengan Prosentase *Missing Value* 5%

k	*Rough Set*			*Fuzzy Rough Set*		
	Jumlah *Rules*	Akurasi	Waktu Komputasi	Jumlah *Rules*	Akurasi	Waktu Komputasi
1	122	81.50%	286.80	169	74.57%	11.42
2	119	88.45%	302.99	165	83.82%	8.74
3	118	78.74%	304.03	160	79.89%	11.70
4	119	81.50%	167.76	149	76.88%	10.90
5	118	78.03%	145.55	171	72.83%	13.05
6	124	85.63%	162.57	172	81.61%	12.12
7	122	86.71%	159.73	158	79.77%	7.66
8	119	85.06%	160.92	161	82.18%	12.01
9	124	86.21%	147.41	149	81.03%	13.23
10	124	91.95%	161.23	166	85.06%	7.16
Rata-Rata	120	84.38%	199.90	162	79.76%	10.80

HASIL DAN PEMBAHASAN

Analisa Hasil

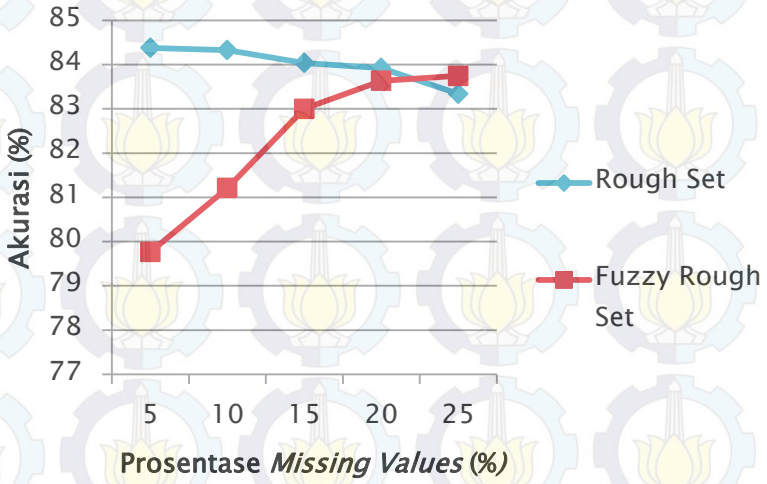
Hasil prediksi oleh *rules* berbasis algoritma *rough set* dan *rules* berbasis algoritma *fuzzy rough set*

SR	SR
SR	SR
SR	SR
SR	SR
SR	SR
-	SR
SR	SR
SR	SR
SR	SR
SR	SR
SR	SR
SR	SR

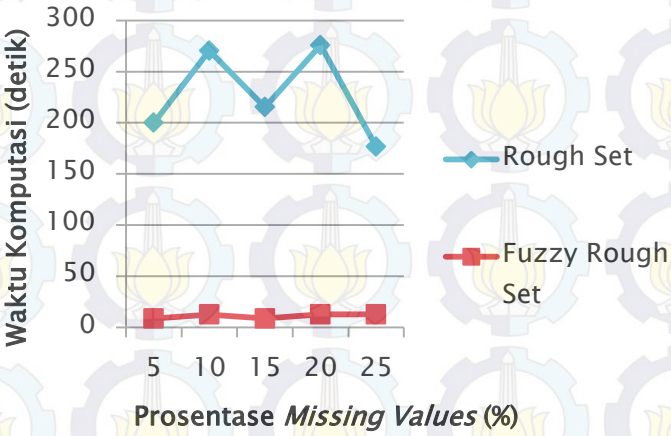
HASIL DAN PEMBAHASAN

Analisa Hasil

Perbandingan akurasi rata-rata *rules* dan waktu komputasi rata-rata dari pembentukan *rules* berbasis algoritma *rough set* dan *fuzzy rough set*



Grafik Perbandingan Akurasi *Rules* Berbasis Algoritma *Rough Set* dan *Fuzzy Rough Set*



Grafik Perbandingan Waktu Komputasi Pembentukan *Rules* Berbasis Algoritma *Rough Set* dan *Fuzzy Rough Set*

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Dari hasil penelitian diperoleh kesimpulan sebagai berikut:

1. Pada penerapan algoritma *rough set, rules* dibentuk berdasarkan *incomplete reduced decision table yang terdiri dari atribut* $A_1, A_2, A_3, A_4, dan\ CH$. Dengan menggunakan *10-fold validation* diperoleh jumlah rata-rata 120 *rules*.
2. Pada penerapan algoritma *fuzzy rough set, missing value* diprediksi menggunakan *lower* dan *upper approximation* sehingga data tidak lengkap menjadi data lengkap. Dari *lower approximation* diperoleh *certain rules* sedangkan *possible rules* diperoleh dari *upper approximation*. Dengan menggunakan *10-fold validation* diperoleh jumlah rata-rata 162 *rules.*
3. Berdasarkan perbandingan akurasi, waktu komputasi, jumlah *rules*, dan kemampuan *rules* untuk memprediksi data uji dapat disimpulkan bahwa performansi dari *rules* berbasis algoritma *fuzzy rough set* lebih baik daripada *rules* berbasis algoritma *rough set*.
4. Penambahan prosentase *missing value* mempengaruhi akurasi *rules* berbasis algoritma *rough set* dan *fuzzy rough set*. Sedangkan waktu komputasi pembentukan *rules* berbasis algoritma *rough set* dan *fuzzy rough set* tidak dipengaruhi oleh penambahan prosentase *missing value* pada dataset.

KESIMPULAN DAN SARAN

5.2 Saran

Berdasarkan penelitian ini, saran untuk penelitian berikutnya adalah:

1. Algoritma *fuzzy rough set* pada penelitian ini dapat digunakan sebagai tahapan *preprocessing* pada dataset tidak lengkap pada penerapan algoritma klasifikasi lainnya.

2. Untuk penelitian yang sejenis, dapat menambahkan atribut cuaca lainnya, serta mempertimbangkan semua variabel linguistik pada data uji dalam himpunan *fuzzy* untuk memprediksi kelas keputusan curah hujan. Penelitian ini juga dapat dilanjutkan dengan memberikan *threshold* untuk pembentukan *rules*.

DAFTAR PUSTAKA

Arifin, Syamsul dan Aisyah, A.S. (2009), "Aplikasi Sisem Logika Fuzzy pada Peramalan Cuaca di Indonesia Kasus : Cuaca Kota Surabaya", *Seminar Nasional Aplikasi Teknologi Prasarana Wilayah 2009*, ISBN 978-979-18342-1-6

Chmielewski, M.R., Gryzmala-Busse, J.W., Peterson, N.W., dan Than, Soe (1993), "The rule induction system LERS – A version for personal computers", *Foundations of Computing and Decision Sciences*, 18, 181–212.

Derrac, Joaquin, Cornelis, Chris, Garcia, Salvador, dan Herrera, Fransisco (2011), "A preliminary Study on the Use of Fuzzy Rough Set based Feature Selection for Improving Evolutionary Instance Selection Algortms", *IWANN 2011, Part I, LNCS 6691,* pp. 174–182. Spinger, Berlin.

Gryzmala-Busse, J.W. (2001), "A Comparison of Several Approaches to Missing Attribute Values in Data Mining", RSCTC 2000, LNAI 2005, pp. 378-385. Spinger, Berlin.

Gryzmala-Busse, J.W. (2004), "Three Approaches to Missing Attribute Values – A Roug Set Perspective", *Workshop on Foundation of Data Mining, associated with the fourth IEEE International Conference on Data Mining,* UK.

Han, Jiawei, Kamber, Micheline, dan Pei, Jian (2012), *Data Mining : Concepts and Techniques Third Edition,* Morgan Kaufmann, USA.

Hong, Tzung-Pei, Tseng, Li-Huei, dan Cien, Been-Chian (2009), "Mining from Incomplete Quantitative by Fuzzy Rough Sets", *Expert Systems with Application* DOI:10.1016/j.eswa.2009.08.002.

Iqbal, Mohammad, Mukhlash, Imam, dan Astuti, H.M (2013), "The Comparison of CBA Algorithm and CBS Algorithm for Meteorological Data Classification", *Information Systems International Conference (ISICO),* 2-4 December 2013.

Jan, Zahoor, Abrar, M., Bashir, Shariq, dan Mirza, Anwar M. (2008), "Seasonal to Inter-annual Climate Prediction Using Data Mining KNN Technique", *IMTIC* 2008, *CCIS* 20, pp. 40-51. Spinger, Berlin.

Kartasapoetra, A.G. (2012), *Klimatologi Pengaruh Iklim terhadap Tanah dan Tanaman*, PT Bumi Aksara, Jakarta.

DAFTAR PUSTAKA

Kryszkiewicz, M. (1998), "Rough Set Approach to Incomplete Information Systems", *Information Science,* Vol . 112, No. 1, pp. 39-49.

Lakitan, Benyamin (2002), *Dasar-dasar Klimatologi*, PT Raja Grafindo Persada, Jakarta.

Maharani, Warih (2008), "Analisis Performansi Algoritma Rough Adaptive Neuro-Fuzzy Inference System", *Seminar Nasional Aplikasi Teknologi Informasi 2008 (SNATI 2008),* Yogyakarta, 21 Juni 2008, ISSN: 1907-5022

Mujiasih, Subekti (2011), "Pemanfaatan Data untuk Prakiraan Cuaca", *Jurnal Meteorologi dan Geofisika* Vol.12 No. 2 – Sepetember 2011: 185-195.

Nandagopal, S., Karthik, S., dan Arunachalam, V.P. (2010), "Mining of Meteorological Data Using Modified Apriori Algorithm", *European Journal of Scientific Research* Vol. 47 No.2 pp. 295-308.

National Council of Applied Economic Research (2010), *Impact Assessment and Economic Benefits of Weather and Marine Services*. Artikel ini dapat didownload di website http://www.ncacr.org.

Neiburger, Morris, Edinger, J.G., Bonner, W.D., dan Purbo, Ardina (1995), terjemahan Ardina Purbo, *Memahami Lingkungan Atmosfir Kita*, ITB, Bandung.

Nofal, "Alaa Al Deen" Mustafa dan Bani-Ahmad, Sulieman (2010), "Classification Based on Association-Rule Mining Techniques: A General Survey and Empirical Comparative Evaluation", *Ubiquitos Computing and Communication Journal* Vol.5 Number 3 pp. 9-17.

Olaiya, Folorunsho dan Adeyemo, Adesesan Barnabas (2012), "Application of Data Mining Techniques in Weather Prediction and Climate Change Studies", *I.J. Information Engineering and Electronic Business* 2012, 1, 51-59. DOI: 10.5815/ijieeb.2012.01.07.

Pusat Bahasa Departemen Pendidikan Nasional RI (2008), *Kamus Besar Bahasa Indonesia*. Artikel ini dapat didownload di website http://bahasa.kemdiknas.go.id/kbbi/index.php.

DAFTAR PUSTAKA

Prasetya, Y.L.D. (2013), *Respon Masyarakat Daera Perbatasan Kalimantan Barat – Serawak (Paloh) Terhadap Peringatan Dini Cuaca Ekstrim BMKG sebagai Langkah Awal untuk Mengurangi Resiko Bencana Hidrometeorologi*. Artikel ini dapat didownload di website https://www.academia.edu/7340278.

Sadiq, A.T, Dualmi, M.G., dan Shaker, A.S. (2013), "Data Missing Solution Using Rough Set Theory and Swarm Intelligence", *International Journal of Advanced Computer Science and Information Technology (IJACSIT)* Vol. 2, No. 3, 2013, Page: 1-16, ISSN: 2296-1739.

Shen, Qiang dan Jensen, Richard (2007), "Rough Sets, Their Extensions and Applications", *International Journal of Automation and Computing* 04(3), July 2007, 217-228 DOI: 10.1007/s1 1633-007-0217-y.

Tay, F.E.H. dan Shen, Lixiang (2002), "Economic and Financial Prediction Using Rough Sets model", *European Journal of Operational Research* 141 (2002) 641 659 PII: S0377-2217(01)00259-4

Tjasyono, Bayong (2004), *Klimatologi*, ITB, Bandung.

Xiao-feng, Hui dan Song-song, Li (2010), "Research on Predicting Stock Price by Using Fuzzy Rough Set", *International Conference on Management Science & Engineering (17 th*) 978-1-4244-81194/10/$26.00, IEEE.

Zadeh, L.A. (1988), "Fuzzy Logic", *IEEE Computer*, 83-93.

Zhao, Liu dan Chang-lu, Qiao (2009), "Research on Drought Forecast Based on Rough Set Theory", *Second International Symposium on Information Science and Engineering*, DOI 10.1109/ISISE.2009.61, IEEE.

TERIMA KASIH