

Aplikasi Neural-Fuzzy pada Regresi Interval untuk Data Time Series

Sri Kusumadewi

Jurusan Teknik Informatika, Universitas Islam Indonesia, Yogyakarta

Jl. Kaliurang K, 14,5 Yogyakarta

(0274) 895287

E-mail : cicie@fti.uii.ac.id

Abstrak

Neural Fuzzy System (NFS) dirancang untuk merealisasikan proses penalaran fuzzy. Pada NFS, bobot-bobot yang terhubung pada jaringan tersebut berhubungan dengan parameter-parameter penalaran fuzzy. Pada penelitian ini, NFS akan diaplikasikan untuk regresi interval dengan data yang bersifat time series. Koefisien-koefisien regresi direpresentasikan dengan menggunakan bilangan fuzzy. Algoritma pembelajaran backpropagation akan digunakan untuk melatih data time series tersebut untuk mendapatkan bobot-bobot akhir, yang nantinya akan digunakan untuk mengestimasi batas bawah dan batas atas interval.

Data yang digunakan dalam penelitian adalah data musiman dalam 4 periode. Proses pembelajaran dilakukan dengan 79 data pelatihan, dan 14 data pengujian, dengan 3 neuron pada lapisan tersembunyi, laju pembelajaran = 1, momentum = 0,9; toleransi error = 10^{-7} , dan maksimum epoch = 10000. Hasil penelitian menunjukkan bahwa MSE untuk data pelatihan sebesar 0,0032 (untuk BPN⁺) dan 0,0037 (untuk BPN). Sedangkan mean interval antara batas bawah dengan nilai aktual untuk data pelatihan sebesar 3,3246 (BPN⁺) dan 3,3834 (BPN), serta untuk data pengujian sebesar 2,2799 (BPN⁺) dan 4,937 (BPN).

Kata Kunci : fuzzy, regresi, backpropagation

1. Pendahuluan

Model regresi terbukti telah memiliki kehandalan untuk penyelesaian masalah interpolasi maupun ekstrapolasi. Suatu nilai tegas y diperoleh sebagai fungsi dari x , $y = f(x)$. tentu saja nilai y bukanlah nilai eksak, dengan kata lain terdapat toleransi error antara nilai y sebagai fungsi dari x , dengan nilai aktual yang diharapkan. Tentu saja toleransi error diharapkan cukup kecil sehingga akurasi y sebagai hasil regresi juga cukup tinggi. Untuk

mengakomodasi hal ini, dapat dibuat suatu interval 'kebolehan' yang mana keberadaan data hasil regresi dalam interval tersebut masih diperbolehkan (mendapatkan toleransi). Interval ini sebenarnya dapat direpresentasikan dengan memberikan nilai koefisien regresi sebagai bilangan fuzzy. Misalkan diberikan pasangan input – output (x_k, d_k) , $k = 1, 2, \dots, p$ dengan $x_k = (x_{k1}, x_{k2}, \dots, x_{kn})$. Model regresi fuzzy pada pola ke- k direpresentasikan sebagai:

$$Y(x_k) = A_0 + A_1x_{k1} + \dots + A_nx_{kn} \quad (1)$$

dengan A_i adalah bilangan fuzzy. Oleh karena itu, nilai estimasi output $Y(x_k)$ juga merupakan bilangan fuzzy.

Beberapa penelitian dan aplikasi tentang regresi interval telah dilakukan dalam Changha (2006), Dug (2005), Dug (2001), Tseng (2002).

Jaringan syaraf dengan algoritma pembelejaraan backpropagation juga telah terbukti dapat digunakan untuk pendekatan fungsi dengan menggunakan proses pembelajaran, sebagai pengganti regresi. Pada penelitian ini, akan digunakan jaringan syaraf backpropagation untuk pendekatan fungsi (regresi) dengan koefisien-koefisien yang direpresentasikan dengan bilangan fuzzy. Adapun data yang akan dilatih dan diuji merupakan data yang bersifat *time series*.

2. Dasar Teori

Konsep dasar dari analisis regresi interval yang didasarkan pada jaringan backpropagation, diperkenalkan oleh Ishibuchi dan Tanaka (1992). Model tersebut menggunakan 2 jaringan backpropagation. Satu jaringan digunakan untuk batas atas interval, sedangkan satu jaringan lainnya digunakan untuk batas bawah interval data. Kedua jaringan tersebut dilatih secara terpisah .

Misalkan $g^+(x_k)$ dan $g^-(x_k)$ adalah output dari kedua jaringan backpropagation tersebut (BPN⁺ dan BPN⁻) yang

berhubungan dengan input vektor x_k , dimana setiap jaringan memiliki n neuron pada lapisan input dan 1 neuron pada lapisan output. Proses pembelajaran dilakukan terhadap kedua jaringan (BPN^+ dan BPN^-) untuk mendapatkan output jaringan $g^+(x_k)$ dan $g^-(x_k)$ yang berkaitan dengan kondisi sebagai berikut:

$$g^-(x) \leq d_k \leq g^+(x), k = 1, 2, \dots, p. \quad (2)$$

Pada proses pembelajaran BPN^+ , fungsi biaya yang digunakan adalah:

$$E = \sum_{k=1}^p E_k = \frac{1}{2} \sum_{k=1}^p \alpha_k [d_k - g^+(x_k)]^2 \quad (3)$$

dengan α_k diberikan sebagai berikut:

$$\alpha_k = \begin{cases} 1; & d_k > g^+(x_k) \\ \alpha; & d_k \leq g^+(x_k) \end{cases} \quad (4)$$

dengan α adalah bilangan positif yang cukup kecil pada interval (0, 1). Nilai α ini dapat diberikan dengan menggunakan fungsi penurunan, sebagai:

$$\alpha(t) = \frac{1}{\left[1 + \left(\frac{t}{2000}\right)\right]^3} \quad (5)$$

dengan t adalah iterasi ke.

Aturan pembelajaran backpropagation dengan 3 lapisan (1 lapisan input, 1 lapisan tersembunyi, dan 1 lapisan output) dan fungsi aktivasi sigmoid, digunakan untuk mendapatkan bobot-bobot w_j (bobot antara lapisan tersembunyi dengan lapisan output) dan w_{ji} (bobot antara lapisan input dengan lapisan tersembunyi) pada jaringan BPN^+ , perbaikan bobot dilakukan sebagai berikut.

$$\Delta w_j = \eta \left(-\frac{\partial E_k}{\partial w_j} \right) = \eta \delta_k y_{kj} \quad (6)$$

$$\Delta w_{ji} = \eta \left(-\frac{\partial E_k}{\partial w_{ji}} \right) = \eta \delta_{kj} x_{ki} \quad (7)$$

dengan

$$\delta_k = \alpha_k (d_k - y_k) y_k (1 - y_k) \quad (8)$$

$$\delta_{kj} = y_{kj} (1 - y_{kj}) \delta_k w_j \quad (9)$$

$y_k = g^+(x_k)$ adalah output jaringan, dan y_{kj} adalah output neuron ke- j pada lapisan tersembunyi untuk input x_k .

Dengan cara yang sama, pembelajaran juga dilakukan pada jaringan BPN^- untuk mendapatkan output jaringan $g^-(x_k)$. Fungsi biaya yang digunakan untuk pembelajaran ini seperti terlihat pada **persamaan (3)**, dengan α_k diberikan sebagai berikut:

$$\alpha_k = \begin{cases} \alpha; & d_k \geq g^-(x_k) \\ 1; & d_k < g^-(x_k) \end{cases} \quad (10)$$

dengan α adalah bilangan positif yang cukup kecil pada interval (0, 1).

Dengan menggunakan kedua algoritma pembelajaran tersebut, kita dapat menentukan 2 fungsi, $g^+(x)$ dan $g^-(x)$ dimana $g^-(x) \leq d_k \leq g^+(x)$, $k = 1, 2, \dots, p$. Dari sini, dapat diperoleh interval:

$$G(x) = [g^-(x), g^+(x)] \quad (11)$$

Karena $g^+(x)$ dan $g^-(x)$ diperoleh dari pembelajaran yang terpisah, maka sangat dimungkinkan $g^+(x) < g^-(x)$. Sehingga, jika hal tersebut terjadi, maka dapat dilakukan modifikasi pada interval tersebut sebagai berikut:

$$h^-(x) = \begin{cases} g^-(x); & g^-(x_k) \leq g^+(x) \\ \frac{1}{2}(g^-(x) + g^+(x)); & g^-(x_k) > g^+(x) \end{cases} \quad (12)$$

$$h^+(x) = \begin{cases} g^+(x); & g^-(x_k) \leq g^+(x) \\ \frac{1}{2}(g^-(x) + g^+(x)); & g^-(x_k) > g^+(x) \end{cases} \quad (13)$$

Dengan demikian, interval yang terjadi adalah:

$$G(x) = [h^-(x), h^+(x)] \quad (14)$$

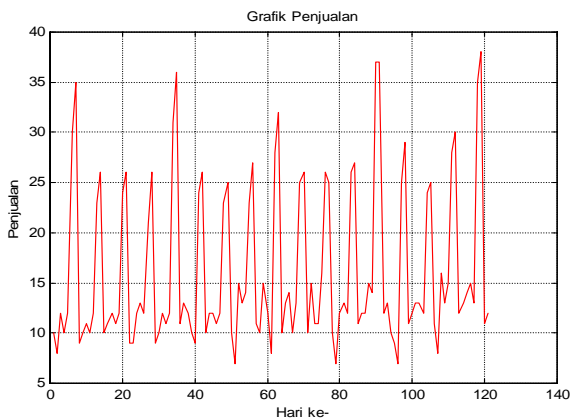
3. Hasil

Diberikan data-data penjualan mie instan dalam 121 hari seperti terlihat pada **Tabel 1**. Kolom hari menunjukkan hari (S=Senin, S=Selasa, R= Rabu, dst), No menunjukkan nomor urut (dari 1 sampai 121), dan P menunjukkan jumlah permintaan mie instan pada hari tersebut dalam satuan dus (Kusumadewi, 2004).

Tabel 1 Data-data penjualan mie instan dalam 4 bulan.

Hari	Tgl	No	P	Tgl	No	P	Tgl	No	P	Tgl	No	P	Tgl	No	P
S	1	1	10	29	29	9	27	57	11	24	85	11	21	113	12
S	2	2	8	30	30	10	28	58	10	25	86	12	22	114	13
R	3	3	12	1	31	12	29	59	15	26	87	12	23	115	14
K	4	4	10	2	32	11	30	60	12	27	88	15	24	116	15
J	5	5	12	3	33	12	31	61	8	28	89	14	25	117	13
S	6	6	30	4	34	31	1	62	28	29	90	37	26	118	35
M	7	7	35	5	35	36	2	63	32	30	91	37	27	119	38
S	8	8	9	6	36	11	3	64	10	31	92	12	28	120	11
S	9	9	10	7	37	13	4	65	13	1	93	13	29	121	12
R	10	10	11	8	38	12	5	66	14	2	94	10			
K	11	11	10	9	39	10	6	67	10	3	95	9			
J	12	12	12	10	40	9	7	68	13	4	96	7			
S	13	13	23	11	41	24	8	69	25	5	97	25			
M	14	14	26	12	42	26	9	70	26	6	98	29			
S	15	15	10	13	43	10	10	71	10	7	99	11			
S	16	16	11	14	44	12	11	72	15	8	100	12			
R	17	17	12	15	45	12	12	73	11	9	101	13			
K	18	18	11	16	46	11	13	74	11	10	102	13			
J	19	19	12	17	47	12	14	75	16	11	103	12			
S	20	20	24	18	48	23	15	76	26	12	104	24			
M	21	21	26	19	49	25	16	77	25	13	105	25			
S	22	22	9	20	50	10	17	78	10	14	106	11			
S	23	23	9	21	51	7	18	79	7	15	107	8			
R	24	24	12	22	52	15	19	80	12	16	108	16			
K	25	25	13	23	53	13	20	81	13	17	109	13			
J	26	26	12	24	54	14	21	82	12	18	110	15			
S	27	27	21	25	55	23	22	83	26	19	111	28			
M	28	28	26	26	56	27	23	84	27	20	112	30			

Data-data tersebut diperlihatkan secara grafis pada **Gambar 1**. Jumlah permintaan akan tinggi pada hari-hari tertentu yaitu Sabtu dan Minggu. Demikian juga, pada awal atau akhir bulan, permintaan mie instan ini bahkan lebih tinggi.



Gambar 1 Data-data penjualan aktual secara grafis.

Input sistem adalah permintaan ke-($t-21$), ($t-14$), ($t-7$) dan (t). Sedangkan output sistem adalah permintaan ke-($t+7$). Dengan demikian akan terdapat rentang data dari 22 hingga 114 (93 data). Dari data tersebut akan digunakan sebanyak 79 data untuk dilatih yang dapat dilihat pada

Tabel 2, sedangkan sisanya sebanyak 14 data akan digunakan sebagai data pengujian yang dapat dilihat pada Tabel 3.

Tabel 2 Tujuh puluh sembilan data yang akan dicluster.

No	Input				Out	No	Input				Out
	t-21	t-14	t-7	t	t+7		t-21	t-14	t-7	t	t+7
1	10	9	10	9	9	41	24	23	23	28	25
2	8	10	11	9	10	42	26	25	27	32	26
3	12	11	12	12	12	43	10	10	11	10	10
4	10	10	11	13	11	44	12	7	10	13	15
5	12	12	12	12	12	45	12	15	15	14	11
6	30	23	24	21	31	46	11	13	12	10	11
7	35	26	26	26	36	47	12	14	8	13	16
8	9	10	9	9	11	48	23	23	28	25	26
9	10	11	9	10	13	49	25	27	32	26	25
10	11	12	12	12	12	50	10	11	10	10	10
11	10	11	13	11	10	51	7	10	13	15	7
12	12	12	12	12	9	52	15	15	14	11	12
13	23	24	21	31	24	53	13	12	10	11	13
14	26	26	26	36	26	54	14	8	13	16	12
15	10	9	9	11	10	55	23	28	25	26	26
16	11	9	10	13	12	56	27	32	26	25	27
17	12	12	12	12	12	57	11	10	10	10	11
18	11	13	11	10	11	58	10	13	15	7	12
19	12	12	12	9	12	59	15	14	11	12	12
20	24	21	31	24	23	60	12	10	11	13	15
21	26	26	36	26	25	61	8	13	16	12	14
22	9	9	11	10	10	62	28	25	26	26	37
23	9	10	13	12	7	63	32	26	25	27	37
24	12	12	12	12	15	64	10	10	10	11	12
25	13	11	10	11	13	65	13	15	7	12	13
26	12	12	9	12	14	66	14	11	12	12	10
27	21	31	24	23	23	67	10	11	13	15	9
28	26	36	26	25	27	68	13	16	12	14	7
29	9	11	10	10	11	69	25	26	26	37	25
30	10	13	12	7	10	70	26	25	27	37	29
31	12	12	12	15	15	71	10	10	11	12	11
32	11	10	11	13	12	72	15	7	12	13	12
33	12	9	12	14	8	73	11	12	12	10	13
34	31	24	23	23	28	74	11	13	15	9	13
35	36	26	25	27	32	75	16	12	14	7	12
36	11	10	10	11	10	76	26	26	37	25	24
37	13	12	7	10	13	77	25	27	37	29	25
38	12	12	15	15	14	78	10	11	12	11	11
39	10	11	13	12	10	79	7	12	13	12	8
40	9	12	14	8	13						

Tabel 3 Empat belas yang akan dicek.

No	Input				Out
	t-21	t-14	t-7	t	t+7
1	12	12	10	13	16
2	13	15	9	13	13
3	12	14	7	12	15
4	26	37	25	24	28
5	27	37	29	25	30
6	11	12	11	11	12
7	12	13	12	8	13
8	12	10	13	16	14
9	15	9	13	13	15
10	14	7	12	15	13

No	Input				Out
	t-21	t-14	t-7	t	t+7
11	37	25	24	28	35
12	37	29	25	30	38
13	12	11	11	12	11
14	13	12	8	13	12

Selanjutnya, pada data pelatihan, dilakukan pembelajaran, masing-masing untuk BPN⁻ dan BPN⁺ menggunakan jaringan backpropagation, dengan 3 neuron pada lapisan tersembunyi, laju pembelajaran = 1, momentum = 0,9; toleransi error = 10⁻⁷, dan maksimum epoch = 10000.

Hasil pembelajaran memberikan nilai h⁻ dan h⁺ untuk 79 data pelatihan seperti terlihat pada Tabel 4. Hasil penelitian menunjukkan bahwa MSE untuk data pelatihan sebesar 0,0032 (untuk BPN⁺) dan 0,0037 (untuk BPN⁻). Sedangkan mean interval antara batas bawah dengan nilai aktual untuk data pelatihan sebesar 3,3246 (BPN⁺).

Tabel 4 Hasil h⁻ dan h⁺ untuk 79 data pelatihan.

Data ke-	h ⁻	y	h ⁺
1	7,4956	9	13,1494
2	7,3544	10	12,7664
3	7,6462	12	14,7899
4	7,4768	11	14,2396
5	7,6559	12	14,9240
6	27,3239	31	37,7235
7	29,9753	36	37,9291
8	7,4466	11	13,0412
9	7,5180	13	13,6891
10	7,5570	12	14,5123
11	7,4533	10	13,7489
12	7,6559	9	14,9240
13	22,0675	24	27,7239
14	25,8766	26	30,6725
15	7,5090	10	13,6824
16	7,5690	12	14,5013
17	7,6559	12	14,9240
18	7,5772	11	14,1773
19	7,6488	12	14,1515
20	22,1800	23	27,7043
21	24,2355	25	28,7445
22	7,4088	10	13,0723
23	7,3798	7	13,6508
24	7,6559	15	14,9240
25	7,7806	13	15,3687
26	7,6998	14	15,2068
27	19,9880	23	22,9741
28	26,3266	27	31,6817
29	7,4300	11	13,3699
30	7,4802	10	13,0569
31	7,6674	15	15,7391
32	7,5588	12	14,5690
33	7,6326	8	15,0564

Data ke-	h^-	y	h^+
34	28,1415	28	37,8184
35	30,2374	32	37,9432
36	7,5756	10	14,1113
37	7,8343	13	15,7061
38	7,6388	14	15,5130
39	7,4532	10	14,0181
40	7,3776	13	12,8688
41	22,7887	25	30,2391
42	25,5625	26	31,6159
43	7,4802	10	13,4524
44	7,6472	15	14,7256
45	7,6777	11	15,6576
46	7,5621	11	14,1132
47	7,7350	16	15,8434
48	21,4213	26	26,1273
49	24,0418	25	27,9887
50	7,5007	10	13,6337
51	7,2703	7	14,1757
52	8,1300	12	16,9587
53	7,7914	13	15,4850
54	7,8633	12	16,6320
55	22,4629	26	26,3358
56	26,7314	27	34,5216
57	7,5765	11	13,8625
58	7,4368	12	12,9283
59	8,1100	12	17,7355
60	7,6538	15	15,0124
61	7,3003	14	13,8620
62	26,8210	37	36,4678
63	29,0776	37	37,8192
64	7,4952	12	13,7569
65	7,8711	13	16,5338
66	7,8875	10	16,1439
67	7,4547	9	14,8557
68	7,8544	7	16,5167
69	25,0292	25	27,9268
70	25,7067	29	30,1135
71	7,4777	11	13,9681
72	7,9680	12	17,1048
73	7,5553	13	13,9793
74	7,5203	13	13,6843
75	8,1522	12	17,2844
76	24,0371	24	28,5256
77	22,9684	25	24,9640
78	7,4678	11	13,7932
79	7,2779	8	13,6335

$$W = \begin{bmatrix} 4,2362 & 1,7860 & 2,7163 & 1,7083 \\ 1,0858 & 0,3289 & -0,1033 & 0,7612 \\ -4,3481 & 0,1017 & 1,6868 & 0,4996 \end{bmatrix}$$

bobot akhir lapisan lapisan tersembunyi ke lapisan output:

$$W = [4,1439 \quad 0,3230 \quad -3,8143];$$

bobot akhir untuk bias pada lapisan input ke lapisan tersembunyi:

$$b = [-4,217; \quad -0,9487; \quad -0,3760];$$

bobot akhir untuk bias pada lapisan tersembunyi ke lapisan output: $b = -3,0207$.

Dari hasil pembelajaran terhadap BPN^+ , diperoleh bobot akhir lapisan lapisan input ke lapisan tersembunyi:

$$W = \begin{bmatrix} 9,8411 & -0,9293 & -1,8426 & -1,2765 \\ 0,1175 & 2,0530 & 0,4241 & 3,5616 \\ 1,0905 & -0,4752 & -1,0018 & -0,4112 \end{bmatrix}$$

bobot akhir lapisan lapisan tersembunyi ke lapisan output:

$$W = [7,3025 \quad 1,9672 \quad 1,2096];$$

bobot akhir untuk bias pada lapisan input ke lapisan tersembunyi:

$$b = [-4,4358; \quad -0,8574; \quad -0,7905];$$

bobot akhir untuk bias pada lapisan tersembunyi ke lapisan output: $b = -2,6993$.

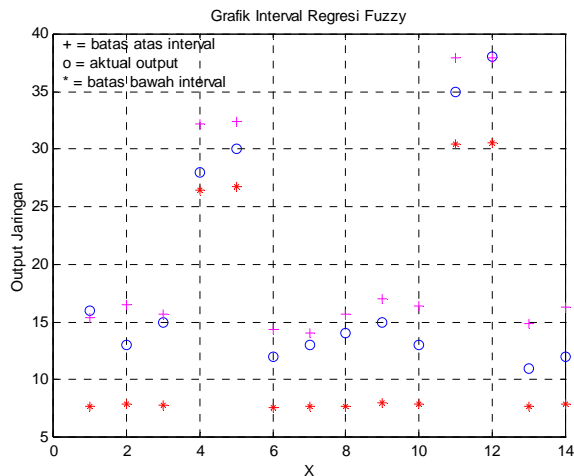
Tabel 5 Hasil h^- dan h^+ untuk 14 data pengujian.

Data ke-	h^-	y	h^+
1	7,6865	16	15,3744
2	7,8524	13	16,5000
3	7,7489	15	15,6855
4	26,4407	28	32,2051
5	26,7945	30	32,4237
6	7,5710	12	14,3063
7	7,6555	13	14,0302
8	7,6373	14	15,6555
9	8,0021	15	17,0405
10	7,8468	13	16,4334
11	30,4305	35	37,9526
12	30,5666	38	37,9471
13	7,6605	11	14,8794
14	7,8268	12	16,2800

Dari hasil pembelajaran terhadap BPN^- , diperoleh bobot akhir lapisan lapisan input ke lapisan tersembunyi:

Selanjutnya, bobot-bobot akhir ini digunakan untuk menghitung h^- dan h^+ untuk data pengujian, hasilnya dapat dilihat pada **Tabel 5**. Sedangkan mean interval antara batas bawah dengan nilai aktual untuk data

pengujian adalah 3,3834 (BPN⁻), serta untuk data pengujian sebesar 2,2799 (BPN⁺) dan 4,937 (BPN⁻).



Gambar 2 Hasil h^- , h^+ , dan output aktual untuk data pengujian.

Gambar hasil h^- , h^+ , dan output aktual untuk data pengujian dapat dilihat pada **Gambar 2**.

4. Kesimpulan

Berdasarkan uraian di atas, dapat disimpulkan bahwa:

1. regresi interval mampu mengakomodasi rentang nilai hasil peramalan pada data yang bersifat *time series*.
2. jaringan syaraf dengan algoritma pembelajaran backpropagation dapat digunakan untuk melatih data pelatihan untuk menentukan batas bawah dan batas atas interval.

Daftar Pustaka

[1]. Changha, Hwang; Dug, Hun Hong; Kyung, Ha Seok. 2006. "Support vector interval regression machine for crisp input and output data". *Fuzzy Sets and Systems* 157(8): 1114-1125 (2006)

[2]. Dug, Hun Hong; Changha; Hwang. 2005. "Interval regression analysis using quadratic loss support vector machine". *IEEE T. Fuzzy Systems* 13(2): 229-237.

[3]. Dug, Hun Hong; Sungho; Lee; Hae, Young Do. 2001. "Fuzzy linear regression analysis for fuzzy input-output data using shape-preserving operations". *Fuzzy Sets and Systems* 122(3): 513-526.

[4]. Kusumadewi, Sri. 2004. *Membangun Jaringan Syaraf Tiruan dengan Tollbox Matlab & ExcelLink*. Graha Ilmu, Yogyakarta.

[5]. Lin, Chin-Teng; dan Lee, George. 1996. *Neural Fuzzy Systems*. London: Prentice-Hall.

[6]. Tseng, Fang-Mei; Gwo-Hshiong; Tzeng. 2002. "A fuzzy seasonal ARIMA model for forecasting". *Fuzzy Sets and Systems*. Volume 126 Pages: 367 – 376.

[CV Penulis]

Sri Kusumadewi, menyelesaikan studi S1 bidang Ilmu Komputer UGM pada tahun 1994, S2 bidang Sistem Komputer & Informatika (Teknik Elektro) UGM pada tahun 1997, dan sekarang sedang menempuh S3 di bidang Ilmu Komputer UGM.

