

## KLASTERISASI PENDERITA DIABETES MENGUNAKAN METODE *Self Organizing Maps (SOM)*

Cevy<sup>1</sup>, Teguh Ikhlas Ramadhan<sup>2</sup>, Agus Supriatman<sup>3</sup>

<sup>1,2,3</sup>Teknik Informatika, Universitas Perjuangan

Jl. Peta No.177, Tasikmalaya, Jawa Barat, (0265) 326058

e-mail: cevy069@gmail.com , teguhikhlas@unper.ac.id , agussupriatman@unper.ac.id

### Abstrak

Klasterisasi yaitu mengelompokkan objek data ke dalam kelompok dengan hubungan paling sedikit dengan kelompok lainnya adalah teknik data mining yang sangat membantu dalam penemuan pengetahuan. Pengelompokan dilakukan berdasarkan rekam medis data RSUD dr.soekardjodan RSUD Ciamis. diabetes mellitus adalah gangguan metabolik yang ditandai oleh kenaikan glukosa darah sebagai hasil dari penurunan sekresi insulin oleh sel beta pankreas atau resistensi insulin. penelitian ini bertujuan untuk menentukan nilai akurasi, pengkelasan dan perbandingan dengan penelitian sebelumnya. Metode ini menggunakan Self Organizing Map (SOM) yang merupakan metode untuk membagi pola masukan ke dalam beberapa kelompok cluster. Pembentukan cluster diawali dengan normalisasi data, kemudian melakukan proses iterasi untuk mendapatkan neuron. Hasil analisis mendapatkan kesalahan kuantisasi 0,45, dan skor siluet 0,50, yang menunjukkan bahwa cluster tersebut kurang kompak, namun hasil silhouette score yang tinggi menunjukkan klaster tersebut terpisah dengan baik dari klaster lainnya.

**Kata kunci**— Klasterisasi, Data Mining, Diabetes Mellitus, Self Organizing Map

### Abstract

*Clustering, which is grouping data objects into groups with the least relationship with other groups, is a data mining technique that is very helpful in knowledge discovery. The grouping was carried out based on medical records from RSUD dr.Soekardjo and RSUD Ciamis. Diabetes mellitus is a metabolic disorder characterized by a rise in blood glucose as a result of decreased insulin secretion by pancreatic beta cells or insulin resistance. This study aims to determine the value of accuracy, classification and comparison with previous research. This method uses Self Organizing Map (SOM) which is a method to divide input patterns into several cluster groups. The formation of clusters begins with normalizing data, then carrying out an iteration process to obtain neurons. The results of the analysis obtained a quantization error of 0.45, and a silhouette score of 0.50, which indicates that the cluster is less compact, but a high silhouette score results indicate that the cluster is well separated from other clusters.*

**Keywords**— Clustering, Data Mining, Diabetes Mellitus, Self Organizing Maps

### 1. Pendahuluan

Menurut Yordan et al. dalam (Dongga et al., 2023), data mining adalah proses yang menggabungkan berbagai metode pengumpulan dan penggunaan data untuk menemukan pola penting dalam data. Dalam proses ini, berbagai metode seperti estimasi, klasifikasi, prediksi, klastering, dan asosiasi digunakan untuk menemukan pola dalam kumpulan data yang sangat besar.

Klasterisasi, yang berarti mengelompokkan objek data ke dalam kelompok dengan hubungan paling sedikit dengan kelompok lainnya, adalah teknik data mining yang sangat membantu dalam penemuan pengetahuan (Yaumi et al., 2020).

Menurut International Diabetes Federation (IDF), diabetes mellitus (DM) adalah gangguan metabolik yang ditandai oleh kenaikan glukosa darah yang disebabkan oleh resistensi insulin atau penurunan sekresi insulin oleh sel beta pancreas (Medika Jurnal et al., 2023).

Tahun 2018, Kota Bandung memiliki 2.849 kasus DM tipe 2, menurut Riset Kesehatan Dasar (RISKESDAS). Pada tahun 2021, diabetes menjadi penyebab kematian nomor satu dengan 1,5 juta insiden, dengan sekitar 727.000 kematian, 48% di antaranya terjadi pada orang di bawah usia 70 tahun.

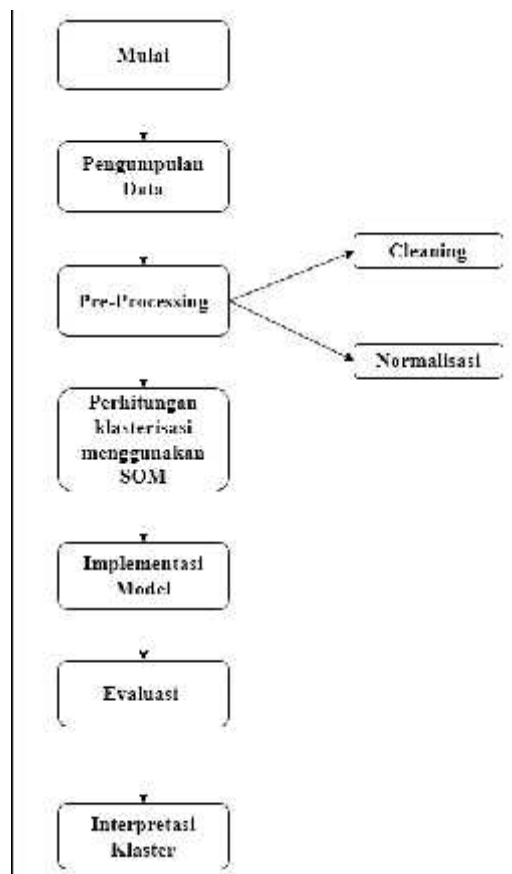
Provinsi Jawa Barat memiliki 52.511 kasus diabetes mellitus, dengan sekitar 26.448 kasus untuk laki-laki dan 26.448 kasus untuk perempuan. Menurut data dari Dinas Kesehatan Kabupaten Ciamis pada semester pertama (Januari hingga Juni 2020), ada 27.483 kasus diabetes mellitus dan 3.352 orang dengan kadar gula rata-rata. Sebagian besar kematian akibat diabetes mellitus terjadi pada usia di bawah 70 tahun. (Amani et al., 2023)

Penelitian sebelumnya tentang klasifikasi diabetes mellitus, yaitu “Prediksi Penderita Diabetes Menggunakan Metode *Naïve Bayes*” (Rahayu, 2023), menghasilkan nilai akurasi sebesar 95,92%. Berdasarkan temuan ini, penelitian selanjutnya akan menggunakan metode *Self Organizing Map (SOM)* dalam kasus yang sama dengan tujuan untuk menentukan nilai akurasi, pengkelasan dan perbandingan dengan penelitian sebelumnya.

Self Organizing Maps (SOM) mengacu pada clustering yang menggunakan data yang hampir identik untuk membentuk satu cluster jika sama; namun, jika data tersebut memiliki karakteristik yang berbeda, maka akan dimasukkan ke dalam cluster yang berbeda. (Fawaz et al., 2022).

**2. Metode Penelitian**

Dalam penelian ini metode kuantitatif adalah yang paling umum digunakan. Tujuannya adalah untuk menentukan seberapa akurat algoritma Self Organizing Maps (SOM) dalam hal penyakit diabetes. Dibawah ini merupakan alur metodologi penelitian :



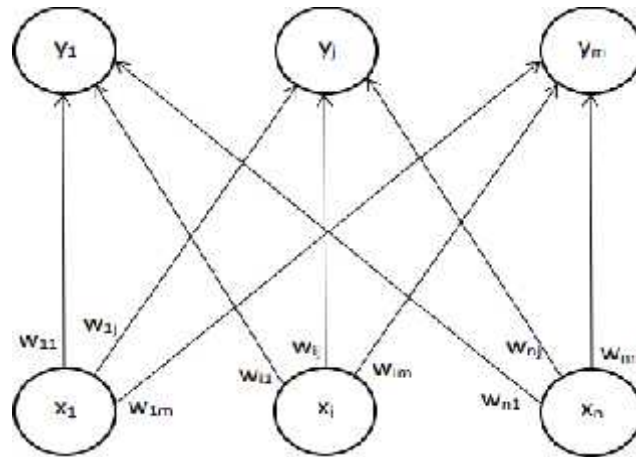
Gambar 1 Metodologi SOM

**Data**

Data atau fakta yang diproses dan diolah menjadi informasi, yang tersusun dengan cara yang mudah dimengerti dan bermanfaat bagi penerimanya. Data dan fakta merupakan “bahan dasar” informasi, tetapi tidak semuanya bisa dijadikan informasi.

**Self Organizing Maps (SOM)**

*Self Organizing Map (SOM)* terdiri dari dua lapisan, yaitu lapisan input dan lapisan output. Setiap neuron pada lapisan input terhubung dengan setiap neuron pada lapisan output, dan setiap neuron pada lapisan output mewakili kelas atau kelompok input yang telah diberikan. Pada Gambar Merupakan arsitektur dari *Self Organizing Map (SOM)*.



Gambar 2 Arsitektur Self Organizing Maps

#### Prosedur Pembentukan SOM

Preprocessing ini dilakukan dengan nilai rata-rata dan normalisasi. Nilai rata-rata tersebut mengubah nilai kosong menjadi bilangan bulat. Normalisasi data adalah salah satu proses yang dilakukan pada fase pra proses data. Pada tahapan ini, mengembalikan nilai-nilai untuk membuat pemrosesan lebih mudah. Dan Normalisasi mengubah data menjadi skala dari nol menit hingga satu maksimum. Selama proses normalisasi data ini, data asli mengalami perubahan linear. Berikut adalah persamaan *Min Max Normalization* ;

$$new_v = \frac{v - \min A}{\max A - \min A}$$

Penjelasan :

$new_v$  : Nilai  $v$  yang baru setelah dinormalisasi

$v$  : Nilai  $v$  yang lama sebelum dinormalisasi

$\max A$  : Nilai maksimum dari variabel  $A$

$\min A$  : Nilai minimum dari variabel  $A$

#### Menentukan Bobot

Bobot akhir akan diperoleh dengan menggunakan colab .

#### Menentukan jarak inter-cluster

*Euclidean Distance* dianggap sebagai distance matrix yang mengadopsi prinsip Phytagoras. Proses perhitungannya menggunakan aturan pangkat dan akar kuadrat. Rumus :

$$D_{(w_i, w_k)} = \sqrt{\sum_{j=1}^p (w_{kj} - x_{ij})^2}$$

Penjelasan :

$D_{(w_i, w_k)}$  : Jarak antara vektor input ke- $i$  dengan bobot neuron ke- $k$

$w_{kj}$  : Bobot neuron ke- $k$  pada variable ke- $j$

$x_{ij}$  : Vektor input ke- $i$  pada variabel ke- $j$

$j$  : Banyak variabel input

Bobot diperbaharui dengan rumus dibawah ini :

$$W_{ij}(\text{baru}) = W_{ij}(\text{lama}) + \alpha * (X_i - W_{ij}(\text{lama}))$$

Penjelasan:

- $W_{ij}(\text{baru})$  : Bobot baru
- $W_{ij}(\text{lama})$  : Bobot lama
- $\alpha$  : *Learning rate*
- $X_i$  : Data hasil normalisasi

### 3. Hasil dan Pembahasan

Pada metode ini pengumpulan data yang diambil dari penelitian terdahulu studi kasus di Rumah Sakit Umum Daerah dr. Soekardjo mendapatkan data sebanyak 502 data dengan 16 atribut pada bulan juli 2019 - Desember 2022. Dan pada penelitian di Rumah Sakit Umum Daerah Ciamis mendapatkan data sebanyak 836 pada bulan Januari 2022 - Desember 2023 berdasarkan tipe gejala penyakit *diabetes mellitus* dari dokumen rekam medis perorangan.

Tabel 1 Data Mentah

No	Umur	Jenis Kelamin	Berat Badan	Sistole	Diastole	GDP	GD2PP	...	Ureum	Kreatinin
1	56	1	64	130	70	212	248	...	36	1.51
2	10	0	25	120	80	171	199	...	8	0.85
3	41	0	62	120	70	223	254	...	31	1.3
...	...	...	...	...	...	...	...	...	...	...
133	56	0	52	145	110	155	184	...	43	1.17

#### Preprocessing Data

Normalisasi data adalah proses mengubah nilai data sebelumnya ke nilai baru. Sebelum melakukan normalisasi, nilai minimum dan maksimum harus dicari. Dalam hal ini, proses normalisasi data dengan mengubah data asli berubah menjadi linear dengan normalisasi minimum dan maksimum.

Tabel 2 Hasil Normalisasi

No	Umur	Jenis Kelamin	Berat Badan	Sistole	Diastole	GDP	GD2PP	...	Ureum	Kreatinin
1	0,66	1	0,313869	0,183	0,1647	0,198	0,25364	...	0,0878	0,00525
2	0,67	1	0,350365	0,178	0,1502	0,099	0,14285	...	0,0139	0,00283
3	0,67	0	0,277372	0,128	0,1676	0,103	0,19970	...	0,0319	0,00241
...	...	...	...	...	...	...	...	...	...	...
13	0,59	0	0,175182	0,057	0,1213	0,101	0,14285	0,051	0,1095	0,41201
38	0361			534	87	243	7	613	19	7

Rumus Normalisasi:

$$new_v = \frac{v - \min A}{\max A - \min A}$$

Nilai minimum GDP = 39

Nilai maksimal GDP = 602

Contoh Perhitungan Normalisasi GDP

1)  $0.198934 = \frac{151-39}{602-39}$

2)  $0.099467 = \frac{95-39}{602-39}$

3)  $0.103019 = \frac{97-39}{602-39}$

4)  $0.113677 = \frac{103-39}{602-39}$

$$5) 0.325044 = \frac{222-39}{602-39}$$

### 3.1. Perhitungan Algoritma *Self Organizing Map (SOM)*

Tabel 3 Data Training

Umur	Jenis Kela min	berat bada n	Sistol e	Diast ole	GDP	GD2P P	...	ureu m	kreati nin
0,662	1	0,313	0,183	0,164	0,198	0,253	...	0,087	0,005
651		869	562	74	934	644		824	25

Tabel 4 Bobot Awal

0	1	2	3	4	5	6	7	8	9	10	11	12	13
0,12	0,41	0,80	0,12	0,55	0,85	0,39	0,05	0,13	0,15	0,83	0,51	0,74	0,98
1343	7998	0433	2915	8821	5722	4485	4223	7619	3044	899	1028	7789	4935

Iterasi 1 data 1 bobot 1

$$D1 = \sqrt{((0,662651 - 0,121343)^2 + (1-0,417998)^2 + (0,313869-0,800433)^2 + (0,183562 -0,122915)^2 + (0,16474 -0,558821)^2 + (0,198934 -0,855722)^2 + (0,253644 -0,394485)^2 + (0,258065 -0,054223)^2 + (0,04913-0,137619)^2 + (0,343348 -0,153044)^2 + (0,401961 -0,83899)^2 + (0,287879 -0,511028)^2 + (0,087824-0,747789)^2 + (0,00525 -0,984935)^2} = 0,293014 + 0,338726 + 0,236744 + 3,678058 + 0,155299 + 0,431370 + 0,01983 + 0,041551 + 7,83030 + 0,036215 + 0,190994 + 0,049795 + 0,435553 + 0,959782 = 13,689431 = \sqrt{13,689431} = 3,7003$$

Bobot diperbaharui :

$$(0,121343, 0,417998, 0,800433, 0,122915, 0,558821, 0,855722, 0,394485, 0,054223, 0,137619, 0,153044, 0,83899, 0,511028, 0,747789, 0,984935) + 0,55 * (0,662651, 1, 0,313869, 0,183562, 0,16474, 0,198934, 0,253644, 0,258065, 0,04913, 0,343348, 0,401961, 0,287879, 0,087824, 0,00525 - 0,121343, 0,417998, 0,800433, 0,122915, 0,558821, 0,855722, 0,394485, 0,054223, 0,137619, 0,153044, 0,83899, 0,511028, 0,747789, 0,984935) = 0,454804, 0,967998, 0,96561, 0,221364, 0,647097, 0,959304, 0,527622, 0,190045, 0,163945, 0,339185, 1,056094, 0,664307, 0,797176, 0,987358$$

Iterasi 1 data 1 bobot 2

$$D2 = \sqrt{((0,662651 - 0,454804)^2 + (1-0,967998)^2 + (0,313869-0,96561)^2 + (0,183562 -0,221364)^2 + (0,16474 -0,647097)^2 + (0,198934 -0,959304)^2 + (0,253644 -0,527622)^2 + (0,258065 -0,190045)^2 + (0,04913-0,163945)^2 + (0,343348 -0,339185)^2 + (0,401961 -1,056094)^2 + (0,287879 -0,664307)^2 + (0,087824-0,797176)^2 + (0,00525 -0,987358)^2} = 0,059579 + 1,024128 + 0,424766 + 1,428991 + 0,232668 + 0,578162 + 0,075063 + 4,626720 + 0,013182 + 1,733056 + 0,427889 + 0,141698 + 0,503180 + 0,964536 = 11,675218 = \sqrt{11,675218} = 3,4178$$

.....

Iterasi 1 data 1 sampai bobot 14

Iterasi 1 data 1 bobot 14

$$D14 = \sqrt{((0,662651 - 1,663994)^2 + (1-2,635599)^2 + (0,313869-1,124537)^2 + (0,183562 -0,50345)^2 + (0,16474 -0,653622)^2 + (0,198934 -0,869343)^2 + (0,253644 -0,795446)^2 + (0,258065 -0,653264)^2 + (0,04913-0,180363)^2 + (0,343348 -0,907998)^2 + (0,401961 -1,360057)^2 + (0,287879 -0,932771)^2 + (0,087824-0,551941)^2 + (0,00525 -0,455361)^2} = 1,002687 + 2,675184 + 0,657182 + 0,102328 + 0,239005 + 0,449448 + 0,293549 + 0,156182 + 0,017222 + 0,318829 + 0,917947 + 0,415886 + 0,215405 + 0,202599 = 7,663453 = \sqrt{7,663453} = 2,768294$$

Perhitungan dilakukan secara manual dengan bobot yang diperbaharui menggunakan jumlah variabel dalam data.

**3.2. Evaluasi Minning Self Organizing Map (SOM)**

Size map, sigma, learning rate, dan iterasi dapat digunakan untuk menghitung nilai *quantization error* dan *silhouette score* setiap eksperimen. Hasilnya dapat dilihat pada Tabel5 dibawah ini.

Tabel 5 Hasil Eksperimen

Percobaan ke	Map Size	Sigma	Learning Rate	Iterasi	Hasil Quantization Error	Hasil Silhouette Score
1	(10, 10)	(0, 3)	(0, 5)	5000	0.33	0.15
2	(12, 12)	(0, 25)	(0, 4)	4000	0.35	0.17
3	(8, 8)	(0, 35)	(0, 6)	3500	0.31	0.15
4	(14, 14)	(0, 22)	(0, 52)	3800	0.36	0.42
5	(9, 9)	(0, 3)	(0, 55)	100	0.45	0.50

Dalam percobaan ketiga, nilai *quantization error* terkecil adalah 0,31, dan skor *silhouette* adalah 0.15. Dalam setiap cluster, titik-titik data berada dekat dengan pusat cluster, yang menunjukkan bahwa cluster tersebut cukup kompak, tetapi skor *silhouette* yang agak rendah menunjukkan bahwa cluster tersebut mungkin tidak terpisah dengan baik dari satu sama lain.

Pada percobaan kelima, kita menemukan kesalahan kuantisasi terbesar, yaitu 0,45 dengan skor *silhouette* 0,50. Kesalahan kuantisasi yang besar menunjukkan bahwa titik data dalam setiap cluster cenderung jauh dari pusat cluster, yang menunjukkan bahwa cluster tersebut kurang kompak. Sementara itu, skor *silhouette* yang tinggi menunjukkan bahwa cluster tersebut terpisah dengan baik dari cluster lainnya.

**3.3. Interpretasi Hasil**

Tabel 6 Hasil klaster

Umur	Jenis Kelamin	berat badan	Sistole	Diasistol	GD P	GD 2PP	HbA1c	Trigliserida (lema)	Kolesterol total	Kolesterol HDL	Kolesterol LDL	Ureum	Kreatinin	Cluster
63	1	60	132	73	151	201	7,7	100	109	54	183	50	1,41	2
64	1	65	130	68	95	125	6,4	93	129	56	202	13	0,78	2
64	0	55	112	74	97	164	6,2	324	152	49	266	22	0,67	1
62	1	64	152	63	103	137	5,8	75	97	41	146	25	0,72	2
52	0	154	133	86	222	36	5,1	140	120	68	223	15	0,83	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
57	0	41	86	58	96	125	4,5	159	125	84	167	19	0,94	1

Hasil dari percobaan kelima menunjukkan bahwa dari 1338 data, klaster 1 mengumpulkan 737 data dan klaster 2 mengumpulkan 601 data, data klaster 1 menunjukkan 737 perempuan dengan umur yang lebih muda dan berat badan yang beragam, meskipun mereka memiliki trigliserida yang tinggi dan tekanan darah sistolik yang berbeda. Namun, kolesterol HDL mereka cenderung baik, yang dapat membantu mencegah penyakit jantung.

Klaster 2, yang terdiri dari 601 data, didominasi oleh laki-laki yang cenderung lebih tua dengan berat badan yang bervariasi, mereka memiliki tekanan darah sistolik yang lebih tinggi, trigliserida dan kolesterol LDL yang lebih tinggi, tetapi mereka juga memiliki kolesterol HDL yang relatif tinggi, yang juga dapat dianggap baik, tetapi cenderung memiliki lebih banyak komplikasi..

**4. Kesimpulan**

Hasil analisis bab sebelumnya tentang klasterisasi penderita diabetes menggunakan metode Self Organizing Map (SOM) pada Rumah Sakit Umum Daerah Kabupaten Ciamis dapat disimpulkan sebagai berikut: ukuran peta 9,9, sigma 0,3, rasio pembelajaran 0,55, iterasi 100, kesalahan quantisasi 0,45, dan skor siluet 0,50. Klaster 1 menunjukkan diabetes tanpa komplikasi, sedangkan Klaster 2 menunjukkan diabetes dengan komplikasi.

Pengujian Map Self Organizing (SOM) menunjukkan akurasi dengan skor silhouette sebesar 0,50 dan quantization error sebesar 0,45, yang menunjukkan bahwa titik data dalam setiap cluster cenderung jauh dari pusat cluster, yang menunjukkan bahwa cluster tersebut kurang kompak. Selain itu, skor silhouette yang tinggi menunjukkan bahwa cluster tersebut terpisah dengan baik dari satu sama lain. Dengan demikian, dapat disimpulkan bahwa metode Self Organizing Map (SOM) cukup efektif untuk mengumpulkan data yang berkaitan dengan diabetes. Metode ini berhasil menemukan pola yang berbeda dalam kumpulan data dan mengelompokkannya menjadi klaster yang relevan..

#### Daftar Pustaka

- [1] Author1A, Author2 B. Title of Manuscript. Name of Journal or its Abbreviation. year; Vol.(Issue): pages.
- [2] Casadei D, Serra G, Tani K. Implementation of a Direct Control Algorithm for Induction Motors Based on Discrete Space Vector Modulation. IEEE Transactions on Power Electronics. 2007; 15(4): 769-777. (in this case Vol.15, Issues 4, and page 769-777)
- [3] Author1 A, Author2 B. Title of Manuscript. Name of Conference of Seminar. City. Year; volume: pages.
- [4] Calero C, Piatini M, Pascual C, Serrano MA. Towards Data Warehouse Quality Metrics. Proceedings of the 3rd Intl. Workshop on Design and Management of Data Warehouses (DMDW). Interlaken. 2009; 39: 2-11. (in this case, city: Interlaken, year: 2009, Vol.39, page: 2-11)
- [5] Author1 A, Author2 B. Title of Manuscript. Name of Conference or Seminar. City. year: pages.
- [6] Yamin L, Wanming C. Implementation of Single Precision Floating Point Square Root on FPGAs. IEEE Symposium on FPGA for Custom Computing Machines. Napa. 2008: 226-232.
- [7] Author1 A, Author2 B. The Title of the Book. Edition. City: Publishert. year: pages.
- [8] Mohan N, Undeland TM, Robbins WP. Power Electronics. New York: John Wiley & Sons. 2005: 11-13.
- [9] Ward J, Peppard J. Strategic planning for Information Systems. Fourth Edition. West Susse: John Willey & Sons Ltd. 2007: 102-104.
- [10] Author1 A, Author2 B. Judul Buku. City: Publisher. Year.
- [11] Mohan N, Undeland TM, Robbins WP. Power Electronics. New York: John Wiley & Sons. 2005.
- [12] Ward J, Peppard J. Strategic planning for Information Systems. Fourth Edition. West Susse: John Willey & Sons Ltd. 2007.
- [13] Author1 A, Author2 B. Editors. Title of the Book. City: Publisher. Year.
- [14] Zade F, Talenta A. Editors. Advanced Fuzzy Control System. Yogyakarta: UAD Press. 2010.
- [15] Author1 A, Author2 B. Title of the Book. In: Editor1 A, Editor2, B. Title of the Book. Edition. City: Publisher. Year: pages.
- [16] Arkanuddin M, Fadlil A, Sutikno T. A Neuro-Fuzzy Control for Robotic Application Based on Microcontroller. In: Krishnan R, Blaabjerg F. Editors. Advanced Control for Industrial Application. 2nd ed. London: Academic Press; 2006: 165-178.
- [17] Originil Author. Year. Title of the Translated Book. Translator. City: Publisher of the translated book. Yearof the translated book.
- [18] Pabla. 2004.Sistem Distribusi Tenaga Listrik. Abdul Hadi. Jakarta: Erlangga. 2007.
- [19] Author. Title of Thesis/Disertation. Thesis/Disertation. City&Name of University/Institute/College; Year.
- [20] Rusdi M. A Novel Fuzzy ARMA Model for Rain Prediction in Surabaya. PhD Thesis. Surabaya: Postgraduate ITS; 2009.
- [21] Author1 A, Author2 B.. Title (this should be in italics). Patent number (Patent). Year of publication.
- [22] Ahmad LP, Hooper A. The Lower Switching Losses Method of Space Vector Modulation. CN103045489 (Patent). 2007.
- [23] Name of Standard Body/Institution. Standard number. Title (this should be in italics). Place of publication. Publisher. Year of publication.

- [24] IEEE Standards Association. 1076.3-2009. IEEE Standard VHDL Synthesis Packages. New York: IEEE Press; 2009.
- [25] Author/Editor (if it is an editor/editors always put (ed./eds.) after the name). Title (this should be in italics). Organisation. Report number: (this should be followed by the actual number in figures). Year of publication.
- [26] James S, Whales D. The Framework of Electronic Government. U.S. Dept. of Information Technology. Report number: 63. 2005..