

Proses Pengelompokan Saraf Menggunakan Jaringan Saraf Tiruan (JST) dengan Algoritme Self-Organizing Maps (SOM)

¹Tantri Lestari, ²Abdul Kudus, ²Sutawanir Darwis

^{1,2}Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung,
Jl. Tamansari No.1 Bandung 40116
email: ¹tantrilestaritantri@yahoo.com

Abstract. Artificial Neural Network Method (ANN) with Self-Organizing Maps (SOM) algorithm is one of the methods using unsupervised learning method. This means that this method does not have a variable Y or without any sample data output. So the SOM algorithm can group the input into output in the form of cluster. The validity index used is the Davies-Bouldin Index (DBI). As an application material, data used PMB UNISBA 2016 obtained from Data Processing Administration UNISBA. The result of clustering shows that clustering of data having the smallest DBI of 1,4463 is the output 3 or group size formed by 3 clusters. The results obtained through the initialization of learning rate (η_0) of 0,9 and decreased learning rate ($\eta(t + 1)$) of 0,04 and iteration as much as 500 times. From the clustering results obtained the average of the centroid of each cluster is 86,9009 (cluster to 1), 83,3775 (cluster to 2), and 80,0384 (cluster to 3).

Keywords : Artificial Neural Network, Self-Organizing Maps, Cluster, Davies-Bouldin Index, Learning Rate.

Abstrak. Metode Jaringan Saraf Tiruan (JST) dengan algoritme *Self-Organizing Maps* (SOM) merupakan salah satu metode yang menggunakan metode pembelajaran tidak terawasi (*unsupervised learning*). Artinya metode ini tidak memiliki variabel Y atau tanpa ada contoh data *output*. Sehingga algoritme SOM dapat mengelompokkan *input* menjadi *output* yang berupa *cluster*. Indeks validitas yang digunakan adalah *Davies-Bouldin Index* (DBI). Sebagai bahan aplikasi, digunakan data PMB UNISBA tahun 2016 yang diperoleh dari Administrasi Pengolahan Data UNISBA. Hasil pengelompokannya menunjukkan bahwa *clustering* terhadap data yang memiliki DBI terkecil sebesar 1,4463 adalah ukuran *output* 3 atau kelompok yang terbentuk sebanyak 3 *cluster*. Hasil tersebut diperoleh melalui inisialisasi *learning rate* (η_0) sebesar 0,9 dan penurunan *learning rate* $\eta(t + 1)$ sebesar 0,04 serta iterasi sebanyak 500 kali. Dari hasil pengelompokan tersebut diperoleh rata-rata dari *centroid* dari masing-masing *cluster* adalah 86,9009 (*cluster* ke-1), 83,3775 (*cluster* ke-2), dan 80,0384 (*cluster* ke-3).

Kata Kunci : Jaringan Saraf Tiruan, Self-Organizing Maps, Cluster, Davies-Bouldin Index, Learning Rate.

A. Pendahuluan

Seleksi mahasiswa baru dalam sebuah perguruan tinggi, umumnya dengan memberikan berbagai jalur seleksi seperti Potensi Minat dan Kemampuan (PMDK) dan Ujian Saringan Masuk (USM). Seleksi tersebut akan dilakukan secara berulang di seluruh perguruan tinggi pada tiap tahunnya. Hal ini dilakukan sebagai proses pencapaian visi dan misi setiap perguruan tinggi untuk mendapatkan berbagai lulusan yang berkualitas. Setiap perguruan tinggi pasti memiliki kualifikasi kelulusan yang berbeda pada tiap tahunnya. Kualifikasi yang biasa digunakan adalah data akademik, data penilaian terhadap sekolah asal, dan data pribadi. Dengan banyaknya kualifikasi tersebut dapat memberikan penilaian yang optimal. Sehingga dapat membuka peluang untuk dihasilkannya informasi yang berguna bagi pihak universitas.

Penggalan informasi pada sebuah data yang berukuran besar (mempunyai jumlah variabel dan jumlah pengamatan yang cukup banyak) tidak dapat dilakukan dengan mudah. *Data mining* merupakan salah satu alat bantu untuk penggalan atau penemuan informasi baru dengan mencari pola atau aturan tertentu pada basis data yang berukuran besar (Davies, et. al., 2004). *Data mining* biasa menggunakan data yang sangat besar. Biasanya data yang besar digunakan untuk membuat hasil lebih dipercaya.

Clustering merupakan suatu kegiatan untuk menemukan kumpulan objek hingga objek-objek dalam satu kelompok sama (atau mempunyai hubungan) dengan yang lain dan berbeda (atau tidak berhubungan) dengan objek dalam kelompok lain. (Fajar Astuti Hermawati, 2013). Dalam metode pembelajaran *data mining*, *clustering* menggunakan metode pembelajaran tak terawasi (*unsupervised learning*). Metode pembelajaran ini tidak memiliki variabel Y, melainkan hanya berdasarkan variabel X. Selanjutnya hasil dari *clustering* ini dapat digunakan untuk memberikan label kelas pada data baru. Salah satu teknik *clustering* dalam *data mining* yang cukup populer adalah algoritme *Self-Organizing Maps* (SOM) dari Jaringan Saraf Tiruan (JST). Algoritme ini merupakan salah satu algoritme yang dapat memberikan hasil *clustering* lebih cocok daripada algoritme *K-Means*. Algoritme SOM juga merupakan algoritme yang kuat dalam mengeksplorasi data dengan *high-dimensional data* dalam aplikasi *data mining*.

Algoritme SOM sudah banyak digunakan oleh peneliti dalam berbagai bidang, salah satunya adalah Edward (2006), telah menggunakan algoritme SOM pada data peserta seleksi mahasiswa baru jalur undangan seleksi masuk Institut Pertanian Bogor (IPB) dengan pilihan pertama program studi di Fakultas Pertanian IPB.

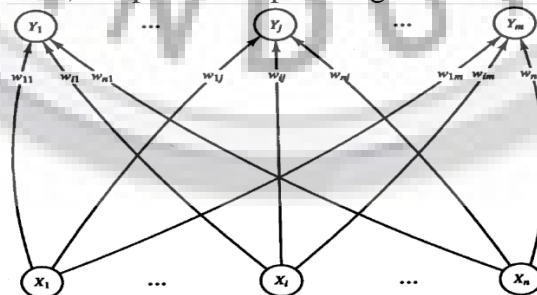
Berdasarkan latar belakang yang telah diuraikan, maka perumusan masalah dalam makalah ini adalah:

1. Bagaimana cara kerja metode JST dengan algoritme SOM?
2. Bagaimana proses pengelompokan menggunakan metode JST dengan algoritme SOM untuk data PMB UNISBA jalur seleksi PMDK tahun 2016?

B. Landasan Teori

Self Organizing Maps (SOM) pertama kali diperkenalkan oleh Prof. Teuvo Kohonen pada tahun 1982. SOM merupakan metode jaringan saraf tiruan yang berbasis *winner takes all*, di mana hanya neuron yang menjadi pemenang yang akan dilakukan pembaruan pada bobotnya. Meskipun berbasis *neural network*, SOM tidak memiliki variabel Y atau tanpa ada contoh data *output*. Sehingga, SOM menggunakan metode pembelajaran tak terawasi (*unsupervised learning*) dengan *competitive-learning rule*.

SOM menggunakan jaringan lapisan tunggal (*single layer network*). Jaringan ini hanya mempunyai satu lapisan bobot koneksi (w_{ij}) dan hanya menerima *input* (x) kemudian secara langsung akan mengolahnya menjadi *output* (y) tanpa harus melalui *hidden layer* (Fausett, 1994). Dapat dilihat pada bagan di bawah ini :



Gambar 1. Arsitektur SOM

Ketika menggunakan SOM, penulis harus melalui berbagai tahapan. Berikut adalah algoritme SOM:

1. Menyiapkan data *input* secara terurut dari $(x_1, \dots, x_i, \dots, x_n)$.
2. Menentukan banyaknya *cluster* yang akan dibentuk.
3. Melakukan inialisasi bobot awal

$(w_{11}, \dots, w_{1j}, \dots, w_{1m}, w_{i1}, \dots, w_{ij}, \dots, w_{im}, \dots, w_{n1}, \dots, w_{nj}, \dots, w_{nm})$
 biasanya menggunakan nilai tengah (*middle point/midpoint*) atau nilai acak dari seluruh data *input*.

4. Menentukan maksimum iterasi.
5. Menentukan laju pembelajaran awal, dimisalkan $\eta_0 = 0,9$.
 Nilai *learning rate* (η) yang digunakan harus dimulai dengan nilai $0 < \eta < 1$.
6. Mencari jarak terdekat melalui nilai bobot w_{ij} dan data *input* menggunakan rumus *euclidean distance*.

$$d_j = \|w_{ij} - x_i\| = \sqrt{\sum_{i=1}^n (w_{ij} - x_i)^2}; i = 1, 2, \dots, n; j = 1, 2, \dots, m. \quad \dots (1)$$

Di mana:

w_{ij} : bobot koneksi antara node *input* ke-*i* terhadap node *output* ke-*j*

x_i : node *input* ke-*i*

n: jumlah node pada lapisan *input*

Langkah ini biasa disebut sebagai *best matching unit*, yaitu menentukan nilai paling cocok dari unit-unit yang memiliki nilai d_j yang minimum. Nilai tersebut merupakan neuron pemenang (*winning neuron*).

7. Memperbarui bobot
 Untuk setiap bobot w_{ij} diperbarui (*update*) vektor bobot pada neuron pemenang dengan persamaan sebagai berikut :

$$w_{ij}(t + 1) = w_{ij}(t) + \eta(x_i - w_{ij}(t)) \quad \dots (2)$$

Di mana:

$w_{ij}(t + 1)$: vektor bobot baru

$w_{ij}(t)$: vektor bobot lama

η : laju pembelajaran (*learning rate*)

8. Memperbarui *learning rate*
 Pembaruan *learning rate* biasa dikatakan sebagai penurunan *learning rate* (PLR). PLR dilakukan setiap kali iterasi, berikut persamaan yang digunakan:

$$\eta(t + 1) = \eta_0 e^{(-\frac{t}{\lambda})} \quad \dots (3)$$

Di mana:

η_0 : *learning rate* awal

t: *current iteration*

λ : jumlah iterasi/banyaknya *cluster* yang dibentuk

Pembaruan *learning rate* ini akan turun secara bertahap tiap kali iterasi, namun tetap di atas 0,01. Sehingga penurunan secara geometrik akan memberikan hasil yang sama.

9. Menyimpan bobot yang telah diperbarui.
10. Mengulangi langkah 6 sampai 9 hingga mencapai kondisi berhenti pada iterasi maksimum yaitu ≥ 500 kali.

Setelah melakukan berbagai tahapan dari algoritme SOM, selanjutnya menghitung validasi hasil *cluster*. Indeks validitas yang digunakan adalah *Davies Bound Index* (DBI). DBI diperkenalkan oleh David dan Donald W. Bouldin (1979) yang digunakan untuk mengevaluasi *cluster*. Validitas internal yang dilakukannya adalah seberapa baik *clustering* sudah dilakukan dengan menghitung kuantitas dan fitur turunan dari set data. *Sum of Square Within Cluster* (SSW) sebagai metrik kohesi dalam sebuah *cluster* ke-*i* diformulasikan oleh persamaan berikut:

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i) \quad \dots (4)$$

m_i adalah jumlah data yang berada dalam *cluster* ke- i , sedangkan c_i adalah sentroid *cluster* ke- i .

Sementara metrik untuk separasi antara dua *cluster* i dan j , digunakan *Sum Of Square Between Cluster* (SSB) dengan mengukur jarak antara sentroid c_i dan c_j , seperti pada persamaan berikut:

$$SSB_i = d(c_i, c_j) \quad \dots (5)$$

Sehingga dapat didefinisikan R_{ij} adalah ukuran rasio seberapa baik nilai perbandingan antara *cluster* ke- i dan *cluster* ke- j . Nilai tersebut diperoleh dari komponen kohesi dan separasi. *Cluster* yang baik adalah yang mempunyai kohesi terkecil dan separasi yang sebesar mungkin. R_{ij} diformulasikan oleh persamaan berikut :

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad \dots (6)$$

Di mana SSW_i adalah rata-rata *error* untuk *cluster* ke- i , SSW_j adalah rata-rata *error* untuk *cluster* ke- j , dan $SSB_{i,j}$ adalah jarak *euclidean* antara *cluster* ke- i dan j .

Sifat-sifat yang dimiliki R_{ij} sebagai berikut :

1. $R_{ij} \geq 0$
2. $R_{i,j} = R_{j,i}$
3. Jika $SSW_j \geq SSW_k$ dan $SSB_{i,j} = SSB_{i,k}$, maka $R_{i,j} > R_{i,k}$
4. Jika $SSW_j = SSW_k$ dan $SSB_{i,j} \leq SSB_{i,k}$, maka $R_{i,j} > R_{i,k}$

Indeks untuk *cluster* ke- i adalah D_i

$$D_i = \max_{j \neq i} R_{i,j} \quad \dots (7)$$

Sehingga, *Davies-Bouldin Index* (DBI) untuk *cluster* ke- i adalah DBI.

$$DBI = \frac{1}{K} \sum_{j=1}^K D_i \quad \dots (8)$$

Di mana K adalah jumlah *cluster* yang digunakan.

Berdasarkan syarat-syarat perhitungan yang didefinisikan di atas, DBI menginginkan nilai sekecil (non-negatif ≥ 0) mungkin untuk menilai baiknya *cluster* yang diperoleh. DBI tidak memiliki batas maksimumnya, sebab tergantung dari proses iterasi yang dilakukan. Semakin banyak iterasi yang dilakukan maka kemungkinan besar DBI akan semakin kecil. Indeks tersebut diperoleh dari rata-rata semua indeks *cluster* dan nilai yang diperoleh bisa digunakan sebagai pendukung keputusan untuk menilai jumlah *cluster* yang paling cocok.

C. Hasil dan Pembahasan

Cara kerja metode JST dengan algoritme SOM terdiri dari berbagai tahapan yang telah dilalui, berikut adalah cara kerjanya:

1. Inisialisasi
 - Menyiapkan data *input* ($x_4, x_5, x_6, \dots, x_{10}$)
 - Banyaknya *cluster* yang akan dibentuk (misal : $j = 1, 2, 3$)
 - Inisialisasi bobot dengan menggunakan *middle point* dari data *input*

Berdasarkan data *input* yang telah ada, dapat dibentuk matriks inisialisasi bobot sebagai berikut:

$$\begin{bmatrix} 82,1 & 77,51 & 70,9 & 79,5 & 81,4 & 81,2 & 81,25 \\ 82,1 & 77,51 & 70,9 & 79,5 & 81,4 & 81,2 & 81,25 \\ 82,1 & 77,51 & 70,9 & 79,5 & 81,4 & 81,2 & 81,25 \end{bmatrix}$$

Inisialisasi *learning rate* (LR) dan penurunan *learning rate* (PLR)

Dalam pembahasan ini digunakan LR sebesar 0,9 (sesuai *default* dalam *software* MATLAB *version* 7.7.0 R2008b) dan PLR sebesar 0,04 yang diperoleh dari

hasil mensubstitusikan $\eta_0 = 0,9$, $\lambda_1 = \frac{500}{3}$, $\lambda_2 = \frac{600}{3}$, $\lambda_3 = \frac{1000}{3}$ melalui rumus pada persamaan 2.3 Bab II sebagai berikut:

$$\begin{aligned} \eta(500) &= 0,9e^{(-500/(500/3))} = 0,0448 \\ \eta(600) &= 0,9e^{(-600/(600/3))} = 0,0448 \\ \eta(1000) &= 0,9e^{(-1000/(1000/3))} = 0,0448 \end{aligned}$$

2. Menghitung jarak terdekat antar neuron dengan rumus *Euclidean distance* sesuai dengan rumus 2.1 pada Bab II.

Misal untuk pengamatan ke-1 (No.Form 9005)

$$d_1 = \sqrt{(82,1 - 80)^2 + (77,51 - 78,2)^2 + \dots + (81,25 - 78,8)^2} = 5,1684$$

$$d_2 = \sqrt{(82,1 - 80)^2 + (77,51 - 78,2)^2 + \dots + (81,25 - 78,8)^2} = 5,1684$$

$$d_3 = \sqrt{(82,1 - 80)^2 + (77,51 - 78,2)^2 + \dots + (81,25 - 78,8)^2} = 5,1684$$

Dari hasil perhitungan jarak yang dilakukan tidak terdapat jarak minimum sebab pada setiap *cluster* diperoleh nilai yang sama. Sehingga dapat diambil secara bebas, dimisalkan diambil neuron pemenang pada *cluster* ke-1.

3. Memperbarui bobot

Pembaruan bobot dilakukan pada *neuron* pemenang. Pembaruan bobot digunakan menurut persamaan 2.2 pada Bab II. Sehingga diperoleh sebagai berikut:

$$w_{11}(baru) = \begin{bmatrix} 82,1 \\ 77,51 \\ 70,9 \\ 79,5 \\ 81,4 \\ 81,2 \\ 81,25 \end{bmatrix} + 0,9 \left(\begin{bmatrix} 80 \\ 78,2 \\ 79,6 \\ 78,8 \\ 84 \\ 82,2 \\ 78,8 \end{bmatrix} - \begin{bmatrix} 82,1 \\ 77,51 \\ 70,9 \\ 79,5 \\ 81,4 \\ 81,2 \\ 81,25 \end{bmatrix} \right) = \begin{bmatrix} 80,21 \\ 78,13 \\ 78,73 \\ 78,87 \\ 83,74 \\ 82,1 \\ 79,05 \end{bmatrix}$$

4. Membentuk pengelompokan berdasarkan jarak minimum. Pada langkah sebelumnya diperoleh neuron pemenang pada pengamatan ke-1 (No.Form 9005) adalah *cluster* ke-1 sehingga No.Form tersebut menduduki *cluster* ke-1.

Algoritme SOM ini diterapkan pada data PMB UNISBA jalur PMDK Tahun 2016. Dengan *input* (rata-rata nilai rapor peserta untuk mata pelajaran PAI, matematika, fisika, kimia, biologi, bahasa indonesia, dan bahasa inggris) dari semester 1-5 berdasarkan latar belakang jurusan sekolah (yang berbasis Sains/Teknologi). Berikut hasil yang telah diperoleh dengan menggunakan bantuan *software* MATLAB 7.7.0 (R2008b).

Tabel 1. Data Hasil Pengelompokan berdasarkan DBI

Banyaknya Cluster	Iterasi		
	500	600	1000
3	1.4460	1.4646	1.7744
4	1.7933	1.8641	1.7806
5	1.8337	1.8928	1.8632
6	2.0154	1.9593	2.3963

Keterangan: LR = 0,9; PLR = 0,04

Berdasarkan Tabel 4.1 di atas, DBI terbaik sebesar 1,4463 dihasilkan dengan tahapan inialisasi (banyaknya *cluster* =3, LR = 0.9, PLR = 0.04 dan 500 iterasi). Dari inialisasi yang digunakan dapat dianalisis bahwa, hampir setiap banyaknya *cluster* akan memberikan peningkatan DBI pada iterasi yang sama, contohnya ketika dilakukan iterasi sebanyak 500 kali. Kemudian, ketika banyaknya iterasi ditambah dan banyaknya *cluster* tetap, maka akan terbentuk DBI yang tidak stabil, contohnya pada

pengelompokan sebanyak 3 *cluster* dan banyaknya iterasi 500, 600, dan 1000. LR dan PLR yang digunakan memiliki nilai yang konstan, sebab agar diperoleh DBI yang optimal (terkecil).

Banyaknya data masing-masing *cluster* dengan pengelompokan sebanyak 3 *cluster* dapat dilihat pada tabel 4.2 dan *centroid* beserta Rata-Rata dari *centroid* dapat dilihat pada Tabel 4.3, serta urutan *cluster* berdasarkan nilai dapat dilihat pada Tabel 4.5 di bawah ini.

Tabel 2. Banyaknya Anggota Masing-Masing *Cluster* dengan 3 *cluster*

<i>Cluster</i> ke-i	Laki-Laki	Perempuan	Banyaknya Anggota	Presentase Banyaknya Anggota
1	165	419	584	30,74%
2	198	483	681	35,84%
3	240	395	635	33,42%
Total	603	1297	1900	100%

Tabel 3. *Centroid* beserta Rata-Rata dari *centroid* dengan 3 *cluster*

<i>Cluster</i> ke-i	<i>Centroid</i>							Rata-rata
	PAI	Matematika	Fisika	Kimia	Biologi	Bahasa Indonesia	Bahasa Inggris	
1	88,8547	86,9727	85,5660	86,4251	86,6393	87,2370	86,6115	86,9009
2	85,7610	82,8210	81,9853	82,2321	82,9287	84,6764	83,2381	83,3775
3	82,3140	78,8300	78,7143	78,6093	79,8976	81,4471	80,4563	80,0384
Rata-Rata	85,6433	82,8745	82,0885	82,4222	83,1552	84,4535	83,4353	83,4389

Tabel 4. Urutan *Cluster* Berdasarkan Nilai

Peringkat ke-i	<i>Cluster</i> ke-i						
	PAI	Matematika	Fisika	Kimia	Biologi	Bahasa Indonesia	Bahasa Inggris
1	1	1	2	2	2	1	2
2	2	3	3	3	3	2	3
3	3	3	3	3	3	3	3

Deskripsi Pengelompokan Terbaik

Cluster ke-1 yang memiliki keanggotaan sebanyak 584 orang yang terdiri dari 165 orang laki-laki dan 419 orang perempuan dengan persentase keanggotaan sebesar 30,74% adalah kelompok terbaik yang memiliki rata-rata nilai rapor tertinggi yaitu 86,9009. *Cluster* ini menduduki peringkat terbaik dan rata-rata nilai rapor tertinggi dari seluruh mata pelajaran yaitu dengan nilai PAI, Matematika dan Bahasa Indonesia.

Cluster ke-2 yang memiliki keanggotaan sebanyak 681 orang yang terdiri dari 198 orang laki-laki dan 483 orang perempuan dengan persentase keanggotaan tertinggi sebesar 35,84% adalah kelompok yang memiliki rata-rata nilai rapor yaitu 83,3775. *Cluster* ini berada di peringkat ke-1 untuk mata pelajaran Fisika, Kimia, Biologi dan Bahasa Inggris, menduduki peringkat ke-2 untuk mata pelajaran PAI dan Bahasa Indonesia, sedangkan berada peringkat ke-3 untuk mata pelajaran Bahasa Indonesia.

Cluster ke-3 yang memiliki keanggotaan sebanyak 635 orang yang terdiri dari 240 orang laki-laki dan 395 orang perempuan dengan persentase keanggotaan sebesar 33,42% adalah kelompok yang memiliki rata-rata nilai rapor tertinggi yaitu 80,0384. *Cluster* ini menduduki peringkat kedua dengan mata pelajaran Matematika, Fisika, Kimia, Biologi, Bahasa Inggris, dan menduduki peringkat ketiga untuk seluruh mata

pelajaran. Namun, untuk mata pelajaran Matematika, Fisika, Kimia, dan Biologi, nilainya berada di bawah rata-rata.

Dari ketiga *cluster* tersebut dapat dilihat bahwa keanggotaan tertinggi pada tiap kelompoknya diraih oleh identitas peserta dengan jenis kelamin perempuan. Hal ini dapat dikatakan bahwa rata-rata nilai rapor jenis kelamin perempuan lebih besar daripada laki-laki pada seleksi PMB UNISBA jalur PMDK tahun 2016 dengan jurusan sekolah asal *Sains/Teknologi*.

Selain dari hasil perhitungan algoritme SOM berdasarkan rata-rata nilai rapor peserta PMB UNISBA tahun 2016, dapat digunakan juga identitas peserta menurut daerah asal yang terdapat pada Tabel 4.6 dan jumlah beserta persentase peserta menurut jurusan asal sekolah (*sains/teknologi*) dapat dilihat pada Tabel 4.7 di bawah ini.

Tabel 5. Persentase Daerah Asal Peserta dalam Setiap *Cluster*

Cluster ke-i	Daerah Asal					
	1	2	3	4	5	6
1	3,79%	24,68%	0,84%	0,37%	1,00%	0,05%
2	3,32%	31,26%	0,32%	0,79%	0,11%	0,05%
3	2,21%	30,63%	0,00%	0,37%	0,16%	0,05%
Rata-rata	3,11%	28,86%	0,39%	0,51%	0,42%	0,05%

Keterangan: 1= Sumatera; 2=Jawa; 3=Nusa Tenggara; 4=Kalimantan; 5=Sulawesi; 6=Papua.

Berdasarkan Tabel 4.6 di atas, dapat diperoleh bahwa dari seluruh *cluster*, peserta asal Sumatera paling banyak di *cluster* 1 (3,79%), dan semakin menurun mengikuti turunnya rata-rata *cluster*. Demikian juga dengan peserta asal Nusa Tenggara (0,84%). Lain halnya dengan peserta asal Jawa, Kalimantan dan Sulawesi, terjadi kenaikan dan penuruna pada tiap *clusternya*. Berbeda dengan peserta asal Papua, memiliki persentase yang konstan pada tiap *clusternya* (0,05%). Secara keseluruhan dengan rata-rata tertinggi adalah peserta asal Jawa yang mendominasi menjadi peserta seleksi PMB UNISBA tahun 2016.

Tabel 6. Jumlah dan Persentase Jurusan Sekolah Asal

Cluster ke-i	Jumlah		Persentase	
	SMA (IPA)	SMK (Sains/Teknologi)	SMA (IPA)	SMK (Sains/Teknologi)
1	568	16	29,89%	0,84%
2	667	14	35,11%	0,74%
3	596	39	31,37%	2,05%
Total	1831	69	96,37%	3,63%

Berdasarkan Tabel 4.7 dapat diperoleh secara keseluruhan bahwa, jumlah dan persentase dari jurusan sekolah asal tertinggi dari tiap *cluster* diduduki oleh SMA jurusan IPA yaitu dengan perolehan sebesar 1829 (96,26%). Sedangkan yang terendah diduduki oleh peserta dengan jurusan asal SMK (*Sains/Teknologi*) dengan perolehan sebesar 71 (3,74%).

D. Kesimpulan dan Saran

Kesimpulan

Terdapat beberapa kesimpulan dalam penulisan makalah ini, antara lain :

1. Metode ANN dengan Algoritme SOM dapat digunakan untuk mengelompokan

data dalam jumlah besar. Prosedur pengelompokannya adalah : (1) menentukan inialisasi (banyaknya *cluster* yang akan dibentuk, bobot awal, *learning rate*, penurunan *learning rate*, dan banyaknya iterasi), (2) menghitung jarak terdekat antar neuron dengan rumus *Euclidean distance*, (3) memperbarui bobot dan *learning rate* pada setiap iterasi yang dilakukan, (4) melakukan pembentukan *cluster* berdasarkan kondisi pemberhentian yang digunakan.

2. Dalam skripsi ini metode ANN dengan algoritme SOM diaplikasikan pada data PMB UNISBA tahun 2016 yang berasal dari jurusan sekolah asal sains/teknologi. Hasil pengelompokan menunjukkan bahwa *clustering* terhadap data yang memiliki DBI terkecil sebesar 1,4463 adalah ukuran *output* 3 atau kelompok yang terbentuk sebanyak 3 *cluster*. Hasil tersebut diperoleh melalui inialisasi *learning rate* sebesar 0,9 dan penurunan *learning rate* sebesar 0,04 dan iterasi sebanyak 500 kali.

Saran

Saran yang dapat dikemukakan dalam penulisan skripsi ini antara lain:

1. Disarankan kepada pihak universitas untuk mempertimbangkan Metode ANN dengan Algoritme SOM untuk pengelompokan data PMB pada tahun selanjutnya, karena dapat digunakan sebagai informasi tambahan untuk mempertimbangkan calon mahasiswa untuk diterima di universitas tersebut.
2. Disarankan kepada peneliti lain dalam melakukan pengelompokan data yang optimal, menggunakan algoritme tambahan seperti *Feed Forward Neural Network*, *Backpropagation*, dan *Support Vektor Machine*. Kemudian dapat membandingkan dengan algoritme lain seperti *K-Means*, *Kernel SOM*, dan *Hierarchical Clustering*. Selain itu, ketika memisalkan ukuran *output* dapat menggunakan analisis faktor agar dapat tercapai DBI yang minimum.

Daftar Pustaka

- Edward. (2006). *Clustering Menggunakan Self Organizing Maps (Studi Kasus: Data PPMB IPB)*. Skripsi. Bogor: Departemen Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Institut Pertanian Bogor.
- Fausett, L. (1994). *Fundamentals of Neural Networks, Architecture, Algorithms, and Applications*. New Jersey: Prentice Hall Inc.
- Kudus, Abdul. (2009). *Decision Tree for Prognostic Classification of Multivariate Survival Data and Competing Risks*. Bandung: Universitas Islam Bandung.
- Guthikonda, Shyam M. (2005). *Kohonen Self-Organizing Maps*. Springfield: Witternberg University.
- Haykin, Simon. (1999). *Neural Network and Learning Machines*. New Jersey: Prentice Hall Inc.
- Mehmed, Kantardzic. (2011). *Data Mining Concepts, Models, Methods and Algorithms*. USA: John Willey & Sons Inc.
- Prasetyo, Eko. (2014). *Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: CV ANDI OFFSET.
- Ridyawan A dan Mukhlason E. (2016). *Pengembangan Sistem rekomendasi Peminjaman Buku Berbasis Web Menggunakan Metode Self Organizing Maps Clustering pada Badan Perpustakaan dan Kearsipan (BAPERSIP) Provinsi Jawa Timur*. Jurnal Teknik ITS, 1, 2301-9271.
- Siang, Jong Jek. (2009). *Jaringan Syaraf Tiruan dan Pemrogramannya Menggunakan Matlab*. Yogyakarta: CV ANDI OFFSET.