

PEMETAAN PREFERENSI MAHASISWA BARU DALAM MEMILIH JURUSAN MENGUNAKAN *ARTIFICIAL NEURAL NETWORK (ANN)* DENGAN *ALGORITMA SELF ORGANIZING MAPS (SOM)*

Muh Najib Hilmi¹, Yuciana Wilandari², Hasbi Yasin³

¹Mahasiswa Jurusan Statistika FSM UNDIP

^{2,3}Staff Pengajar Jurusan Statistika FSM UNDIP

Abstract

College is the highest educational institution and the role the intellectual life of the Indonesian people that the main purpose of academics. Not all colleges into their destination but only college that has a role, credibility and rank the best course of which it is their goal. This makes higher education marketing research approach to get attention and become the main goal of the academics in choosing a college. This research was conducted in order to determine with certainty attribute / emotional reasons academics in choosing college as their academic goals. The method used in this study were self-organizing maps with the Kohonen algorithm is a classification method. Kohonen SOM algorithm with learning rate used 0:05, 0.25, 0:50, 0.75, 0.95 and initialization of initial weight value and the value of the midpoint and 500 iterations with output 3 clusters are formed. Results clustering of SOM validated using Davies-Bouldin index with the best clustering results that DBI minimum (1.7802) with the learning rate is 0.95 and the cluster formed three clusters for the first cluster as many as six members, cluster-2 by 9 members and 3rd cluster as 5 members the results of clustering with top priority contained in the cluster to-2 with a mean (7.434) with the characteristics of each member is an emotional reason in choosing a major.

Keywords: Self Organizing Maps, Kohonen algorithm, Learning Rate, Index Davies Bouldin, Cluster.

1. PENDAHULUAN

Universitas merupakan Institusi Pendidikan Tertinggi yang memberikan gelar akademik dalam berbagai bidang keilmuan. Sebagai institusi tertinggi universitas menjadi tujuan para pelajar, namun tidak semua universitas menjadi tujuan, melainkan universitas yang memiliki predikat baik di tingkat nasional dan internasional yang menjadi tujuan para pelajar. Metode pemasaran telah dilakukan oleh berbagai universitas untuk menjadikan universitas yang bersangkutan menjadi tujuan para pelajar, salah satunya dengan mengetahui karakteristik dari para pelajar yang ingin masuk universitas. Salah satu metode yang dapat digunakan untuk mengetahui hal tersebut adalah ANN (*Artificial Neural Network*). ANN merupakan sistem komputasi yang arsitektur dan operasinya diilhami dari pengetahuan tentang sel syaraf biologis di dalam otak. ANN dapat digambarkan sebagai model matematis dan komputasi untuk fungsi aproksimasi non-linear, klasifikasi data cluster dan regresi non-parametrik atau sebuah simulasi dari koleksi model jaringan syaraf biologi. Salah satu algoritma ANN adalah *Self-Organizing Map (SOM)*. Tujuan dari penelitian ini adalah menjelaskan secara pasti tentang atribut ataupun karakteristik dari para pelajar yang ingin masuk universitas dengan cara mengelompokkan alasan-alasan dan motivasi para pelajar dalam memilih sebuah jurusan.

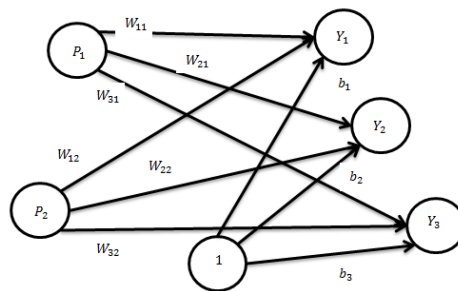
2. TINJAUAN PUSTAKA

Menurut Fauset (1994) dalam Warsito (2009) *Artificial Neural Network (ANN)* atau yang sering disebut dengan *Jaringan Syaraf Tiruan (JST)* merupakan sistem pemroses

informasi yang memiliki karakteristik mirip dengan jaringan syaraf pada makhluk hidup. Adapun metode-metode yang digunakan dalam ANN untuk memecahkan permasalahan cluster diantaranya adalah metode *Kohonen Neural Network*, *Self-Organizing Map*, *Learning Vector Quantization*, *Multilayer Perceptron*, *Support Vector Machine* dan masih banyak lagi metode-metode lainnya (Prasetyo, 2012).

2.1. Self-Organizing Map (SOM)

Self-Organizing Map (SOM) pertama kali diperkenalkan oleh Kohonen dengan teknik pelatihan ANN, yang menggunakan basis *winner takes all*, dimana hanya neuron yang menjadi pemenang saja yang akan diperbarui bobotnya (Prasetyo, 2012). SOM tidak menggunakan nilai target kelas atau yang sering disebut *unsupervised learning* jadi tidak ada kelas yang ditetapkan untuk setiap data. *Self-organizing map* kadang-kadang disebut sebagai Kohonen maps dan salah satu algoritma *self organizing maps* adalah algoritma kohonen. Adapun arsitektur jaringan *self organizing* terlihat pada Gambar 1. Gambar tersebut menggambarkan terdapat 2 unit pada lapisan input (P_1 dan P_2) dan 3 unit pada lapisan output (Y_1 , Y_2 dan Y_3). Pada gambar tersebut terdapat simbol w_{ij} , w_{ij} sendiri adalah sebuah bobot, bobot w_{ij} artinya bobot yang menghubungkan neuron ke-j pada lapisan input ke neuron ke-i pada lapisan output.



Gambar 1. Arsitektur Jaringan *Self Organizing Map*

Salah satu algoritma pembelajaran untuk *SOM* adalah algoritma pembelajaran kompetitif dengan metode Kohonen. Menurut Kusumadewi (2004) algoritma dari *self organizing map* secara berurutan adalah sebagai berikut :

1. a. Inisialisasi bobot input (W_{ij}):

$$W_{ij} = \frac{\text{Min } P_j + \text{Max } P_j}{2} \quad (1)$$

dengan W_{ij} adalah bobot antara variabel input ke-j dengan neuron pada kelas ke-i ($j=1,2,\dots,m; i=1,2,\dots,K$); $\text{Min } P_j$ dan $\text{Max } P_j$ masing-masing adalah nilai terkecil pada variabel input ke-j, dan nilai terbesar dari variabel input ke-j.

- b. Inisialisasi bobot bias (b_i)

$$b_i = e^{[1 - \ln(\frac{1}{K})]} \quad (2)$$

dengan b_i adalah bobot bias ke-neuron ke-i, dan K adalah jumlah cluster.

- c. Set parameter-parameter *learning rate*.
- d. Set maksimum epoh (MaxEpoh)
2. Set Epoh=0;
3. Mengerjakan bila epoh < MaxEpoh:
 - a. Epoh=epoh+1

- b. Pilih data secara acak, misalkan data terpilih adalah data ke-z.
 c. Cari jarak antara data ke-z dengan setiap bobot input ke-i (D_i):

$$D_i = \sqrt{\sum_{j=1}^m (W_{ij} - P_{zj})^2} \quad (3)$$

- d. Hitung $a(i)$: $a_i = -D_i + b_i$
 e. Cari a_i terbesar :
 (i) $\text{MaxA} = \max(a_i)$, dengan $i=1,2,\dots,K$
 (ii) $\text{idx} = i$, sedemikian hingga $a_i = \text{MaxA}$.
 f. Set output neuron ke-i (y_i):

$$y_i = \begin{cases} 1, & i = \text{idx} \\ 0, & i \neq \text{idx} \end{cases}$$

- g. Update bobot yang menuju ke neuron idx :

$$W_{\text{idx}-j} = W_{\text{idx}-j} + \alpha(P_{zj} - W_{\text{idx}-j}) \quad (4)$$

- h. Update bobot bias:

$$(i) \quad c(i) = (1-\alpha)e^{(1-\ln(b(i)))} + \alpha a(i) \quad (5)$$

$$(ii) \quad b(i) = e^{(1-\ln(c(i)))} \quad (6)$$

2.2. Indeks Davies-Bouldin (DBI).

Pendekatan pengukuran ini untuk memaksimalkan jarak *inter-cluster* diantara *Cluster* C_i dan C_l dan pada waktu yang sama mencoba untuk meminimalkan jarak antara titik dalam sebuah *cluster*. Jarak *intra-cluster* $S_c(Q_i)$ dalam *Cluster* Q_i ialah

$$S_c(Q_i) = \left\{ \frac{1}{n_i} \sum_{j=1}^m \sum_{z=1}^{n_i} |X_{ijz} - C_{ij}|^2 \right\}^{\frac{1}{2}} \quad (7)$$

dengan n_i adalah banyak titik yang termasuk dalam *Cluster* Q_i dan C_{ij} adalah *centroid* dari *Cluster* Q_i dari input ke-j dengan $i=1,2,\dots,k$, $j=1,2,3,\dots,m$.

Jarak *Inter-cluster* didefinisikan sebagai.

$$d_{il} = \|C_{ij} - C_{lj}\| \quad (8)$$

dengan C_i dan C_l ialah *centroid Cluster* ke-i dan *Cluster* ke-l dengan $i=1,2,\dots,k$ dan $l=1,2,\dots,k$. Indeks Davies-Bouldin didefinisikan sebagai

$$DB(K) = \frac{1}{K} \sum_{i=1}^K \max \{R_{il}\} \quad (9)$$

dengan nilai $R_{il} = \frac{S_c(Q_i) + S_c(Q_l)}{d_{il}}$, K ialah banyak cluster.

Skema clustering yang optimal menurut indexs Davies-Bouldin adalah yang memiliki Indeks Davies-Bouldin minimal (Davies and Bouldin, 1997).

3. METODE PENELITIAN

Jenis data yang digunakan dalam penelitian ini adalah data primer, yang didapatkan dengan cara membagikan kuisoner kepada mahasiswa angkatan 2013 Fakultas Sains dan Matematika Universitas Diponegoro Semarang. Kuesioner tersebut berisikan alasan-alasan yang mendasari mahasiswa menentukan jurusan yang ditempuh sekarang. Populasi adalah mahasiswa S1 Fakultas Sains dan Matematika Universitas Diponegoro angkatan 2013 yang berjumlah 678 mahasiswa, sedangkan untuk sampel adalah 152 mahasiswa terdiri dari berbagai jurusan Matematika, Statistika, Teknik Informatika, Biologi, Fisika dan Kimia. Pengambilan sampel dengan cara *probability sampling* dengan menggunakan metode *Proporsional Random Sampling*.

Tahap-tahap yang dilakukan pada penelitian ini adalah sebagai berikut :

1. Menentukan variabel yang akan diamati.
2. Membuat rancangan penelitian yang meliputi metode Kohonen *Self Organizing Map*, sumber data, populasi dan sampel penelitian, dan metode pengambilan sampel serta melakukan pengolahan data dengan menggunakan MATLAB 8.0.
3. Melakukan analisis data yang terdiri dari :
 - a. Memasukan data (Input)
 - b. Menentukan variabel-variabel masukan.
 - c. Penyusunan dan Penulisan Coding Algoritma *Self Organizing Map*.
 - d. Pengolahan data menggunakan MATLAB :
 - (a) Randomisasi data sampel masukan.
 - (b) Proses pelatihan/iterasi-iterasi dengan *Self-Organizing Map*.
 - (c) Proses simulasi/bangkitan data masukan dengan SOM
 - (d) Klasifikasi hasil perhitungan.
 - e. Menginteprestasikan hasil output.
 - f. Pengambilan kesimpulan

Variabel – variabel Penelitian adalah :

1. Peluang Karir .
2. Departemen dan Instansi Terbaik untuk bekerja.
3. Peringkat Universitas
4. Akreditasi Jurusan
5. Tenaga Pendidik Profesional.
6. Pelajaran yang difavoritkan di SMA
7. Jumlah alumnus SMA pada jurusan tersebut.
8. Rekomendasi Orang Tua
9. Mengikuti Trand
10. SPP yang terjangkau
11. Biaya Hidup Rendah
12. Jumlah Beasiswa
13. Beasiswa keharusan menempuh jurusan tersebut.
14. Peluang masuk jurusan besar
15. Pilihan alternative
16. Keterlambatan pendaftaran
17. Jarak yang dekat.
18. Keinginan berpetualangan.
19. Terinspirasi Seseorang
20. Keinginan menggapai cita-cita.

4. HASIL DAN PEMBAHASAN

4.1. Hasil Proses dan Perhitungan *SOM Training*

Perhitungan dengan *Self Computing* dengan Menggunakan *learning rate* : 0,05 dengan 500 iterasi dan jumlah *Cluster* yang diinginkan 3 cluster, dengan tujuan mengklasifikasikan masing-masing *cluster* dalam 3 tingkat kepentingan dengan melihat nilai *mean* dari masing-masing cluster. Menggunakan Matlab R2013b didapatkan :

b. Bobot Awal Input *Weight Positions*

Untuk mendapatkan inisialisasi bobot-bobot input pada titik tengah dari minimum dan maksimum input dilakukan perhitungan dengan menggunakan persamaan (1)

Inisialisasi bobot input(W_{ij}): $W_{ij} = \frac{\text{Min } P_j + \text{Max } P_j}{2}$

$$W_{11} = \frac{1+10}{2} = 5,500 \qquad W_{12} = \frac{1+9}{2} = 5,000$$

$$W_{1151} = \frac{5+7,5}{2} = 7,2500 \qquad W_{1152} = \frac{1+10}{2} = 5,500$$

c. Bobot Awal Bias

Adapun hasil dari inisialisasi bobot bias awal dengan menggunakan persamaan (2)

Inisialisasi bobot bias (b_i)

$$b_i = e^{[1-\ln(\frac{1}{K})]} = e^{[1-\ln(\frac{1}{3})]} = 8,1548$$

d. Bobot Akhir Input *Weight*

Setelah program melakukan *Train* Pelatihan Jaringan atau perhitungan iterasi sebanyak 500 iterasi pastinya didapatkan bobot dari input berubah, untuk melihat perubahan bobot input tersebut dapat dilihat pada output dari Matlab yang keluar. Untuk mengetahui perhitungan secara manual maka dihitung dengan persamaan (4) dengan iterasi ke-1 dan ke - 2. Untuk itersi selanjutnya dikerjakan menggunakan komputer.

Perhitungan manual pada iterasi ke-1

$$W_{idx-1} = 5,500 + 0,05(8 - 5,500) = 5.625$$

$$W_{idx-2} = 5,000 + 0,05(7 - 5,000) = 5.1$$

$$W_{idx-151} = 7,250 + 0,05(9 - 7,250) = 7.3375$$

$$W_{idx-152} = 5,000 + 0,05(6 - 5,000) = 5.525$$

e. Bobot Akhir Bias

Adapun hasil dari inisialisasi Bobot Akhir Bias dari Pelatihan jaringan Kohonen ini dengan perhitungan matematika dengan persamaan (5) dan (6).

Update bobot bias:

Untuk perhitungan matematika pada iterasi ke-1 adalah :

(i) $c(1) = (1-0,05)e^{(1-\ln(8,15))} + (0,05 * (-20.4375)) = 0.70521$

(ii) $b(1) = e^{(1-\ln(0.70521))} = 3.854571$

Berikut merupakan hasil perhitungan secara komputer setelah 500 iterasi :

Bobot_Akhir_Bias =

6.0387
9.0625
10.8770

Sedangkan output *cluster* yang terbentuk adalah :

a =

Pertanyaan ke-1 sampai ke-20

1 1 1 1 1 1 0 0 0 0 0 1 0 0 0 0 0 1 0 1
0 0 0 0 0 0 1 0 1 1 0 0 1 0 1 1 0 0 0 0
0 0 0 0 0 0 0 1 0 0 1 0 0 1 0 0 1 0 1 0

ac =

Pertanyaan ke-1 sampai ke-20

1 1 1 1 1 1 2 3 2 2 3 1 2 3 2 2 3 1 3 1

Berdasarkan output dengan Matlab R2013b terlihat secara jelas *Clustering* yang terbentuk yaitu untuk kolom 1 sampai kolom ke-20 ada 3 *Cluster*. Pada kolom urutan ke-1 sampai urutan ke-6 ditunjukkan nilai 1 artinya kolom dari 1 sampai ke-6 menjadi anggota *cluster* ke-1 dan juga kolom ke-12 dan kolom ke-20 sedangkan kolom ke-8, ke-11, ke-14, ke-17, dan kolom ke-19 tertulis angka 3 artinya data-data pada kolom tersebut masuk dalam anggota *Cluster* ke-3 sedangkan kolom yang tersisa masuk dalam *Cluster* ke-2 yang beranggotakan kolom ke-7, ke-9, ke-13, ke-15 dan kolom ke-16.

4.2 Indeks Davies Bouldin (DBI)

Untuk mengetahui *cluster* terbaik diperlukan pengamatan lanjut yaitu pengamatan terhadap DBI, pengamatan ini dilakukan untuk mengukur tingkat validitas dari hasil *clustering*. Parameter penambahan *learning rate* (LR) (α) akan berpengaruh terhadap DBI yaitu semakin besar ataupun sebaliknya. Untuk mengetahui klasifikasi terbaik dengan cara melihat nilai DBI terkecil yang didapatkan. Hasil lengkap perhitungan nilai DBI selengkapnya dengan menggunakan program R. Selain menggunakan *learning rate* 0,05 dicoba menggunakan *learning rate* yang lain dan hasilnya pada Tabel. 1.

Tabel 1. Keanggotan dan nilai DBI masing-masing *Cluster* Menurut *Learning rate*

Data Pertanyaan	Keanggotaan Masing-masing <i>Cluster</i> Menurut <i>Learning rate</i> (α)				
	0,05	0,25	0,50	0,75	0,95
1	1	2	2	2	2
2	1	2	2	2	2
3	1	2	2	2	2
4	1	2	2	2	2
5	1	2	2	2	2
6	1	2	2	2	2
7	2	1	1	1	3
8	3	3	3	3	1
9	2	1	1	1	3
10	2	1	1	1	3
11	3	3	3	3	1
12	1	2	2	2	2
13	2	1	1	1	3
14	3	3	3	3	1
15	2	1	1	1	1
16	2	1	1	1	3
17	3	3	3	3	1
18	1	2	2	2	2
19	3	3	3	3	1
20	1	2	2	2	2
DBI	1.92928	1.92928	1.92928	1.92928	1.780291

Dari pengamatan perhitungan DBI didapatkan nilai DBI terkecil terletak pada *Learning rate* sebesar 0,95 dengan nilai DBI sebesar 1,78 sedangkan yang lain menunjukkan nilai DBI yang lebih besar. Dari hasil *clustering* dengan *Learning rate* sebesar 0,95 diperoleh keanggotan *Cluster* yaitu pertanyaan no.1

sampai ke-6 menjadi anggota *cluster* ke-2 dan juga pertanyaan ke-12, ke-18 dan pertanyaan ke-20 sedangkan pertanyaan ke-8, ke-11, ke-14, ke-15, ke-17, dan pertanyaan ke-19 masuk dalam anggota *Cluster* ke-1 pertanyaan yang tersisa masuk dalam *Cluster* ke-3 yang beranggotakan pertanyaan ke-7, ke-9, ke 10, ke-13 dan pertanyaan ke-16. Interpretasi hasil dari Clustering dengan cara mendapatkan *Cluter* dengan prioritas utama melalui pencarian nilai *mean* dari masing-masing anggota *Cluster*. Untuk *Cluster* dengan prioritas utama ditunjukkan dengan *mean* yang memiliki nilai tertinggi. Hal ini menunjukkan bahwa pada *Cluster* tersebut terdiri dari alasan-alasan terbaik yang mendapat rating tertinggi dari para responden, artinya alasan-alasan tersebut sangat dipertimbangkan dan diperhitungkan oleh seorang mahasiswa dalam memilih jurusannya dan berusaha untuk mendapatkannya. Kebalikannya untuk *Cluster* dengan prioritas bukan utama ditunjukkan dengan nilai *mean* paling rendah dari anggota-anggota *Cluster* yang ada. Artinya anggota dari *Cluster* tersebut berupa alasan-alasan dengan rating yang sangat rendah dan memberikan petunjuk bahwa alasan-alasan tersebut tidak begitu dipertimbangkan oleh responden dalam memilih sebuah jurusan yang diinginkan. Kemudian untuk *Cluster* dengan nilai *mean* yang berada diantara terbaik dan terburuk ditetapkan sebagai *Cluster* penting dari *Cluster* tersebut. Memberikan arti bahwa alasan-alasan yang menjadi anggota *Cluster* tersebut mempunyai reating sedang yang diperoleh dari responden sehingga dilihat dari tingkat kepentingan untuk alasan-alasan tersebut dianggap sedang jadi tidak begitu diperhitungkan oleh responden namun mempengaruhi penilaian mereka dalam memilih sebuah jurusan.

Untuk lebih jelas mengenai klasifikasi tingkat kepentingan dari masing-masing anggota cluster yang terbaik dan juga nilai *mean* dari tiap pertanyaan dapat dilihat pada Tabel 2.

Tabel 2. Hasil Klasifikasi dan Tingkat Kepentingan untuk masing-masing Cluster.

Tingkat Kepentingan	No. Cluster	Nilai Mean	Anggota (no. item motivasi/faktor-faktor)	No. Pertanyaan	Nilai mean setiap pertanyaan	Jumlah Keanggotaan
Sangat Penting	2	7.434	1,2,3,4,5,6,12,18,20	1	8.214079	9
				4	7.731579	
				3	7.608553	
				2	7.496053	
				20	7.437171	
				5	7.422105	
				6	7.257829	
				12	7.117763	
Penting	1	5.288	8, 11, 14, 15, 17, 19	18	6.624342	6
				19	5.768421	
				17	5.529605	
				14	5.390132	
				8	5.316447	
				15	4.992763	
Tidak penting	3	3.800	7, 9,10,13,16	11	4.730921	5
				10	4.570395	
				7	4.269737	
				13	3.855921	
				9	3.496711	
				16	2.811184	

5. KESIMPULAN

Berdasarkan metode clustering *SOM* menggunakan algoritma kohonen dengan *learning rate* yang digunakan 0.05, 0.25, 0.50, 0.75, 0.95 serta 500 iterasi dengan keluaran 3 *cluster* yang terbentuk. Diperoleh *Indeks Davies-Bouldin* dengan hasil *clustering* terbaik yaitu DBI minimal 1.7802 dengan *learning rate* 0.95 dan *cluster* yang terbentuk 3 *cluster* untuk *cluster* pertama sebanyak 6 anggota, *cluster* ke-2 sebanyak 9 anggota dan *cluster* ke-3 sebanyak 5 anggota. Hasil *cluster* dengan prioritas utama terdapat pada *cluster* ke-2 dengan nilai mean 7,434 dengan karakteristik masing-masing anggota adalah alasan emosional dalam memilih sebuah jurusan. Alasan tersebut diantaranya propek pekerjaan dari sebuah jurusan, instansi-intansi peting dalam pemerintahan dan swasta, popularitas dan peringkat Undip, akreditasi jurusan, tenaga pendidik yang profesional, pelajaran di SMA yang difavoritkan berkaitan dengan jurusan, jumlah ketersediaan beasiswa pada jurusan, ingin mencari pengalaman baru dan meraih cita-cita yang diinginkan.

6. DAFTAR PUSTAKA

- Davies, DL dan Bouldin, D.W. 1997. "A cluster separation measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 1, no. 2, pp.224-227 <http://dl.acm.org/citation.cfm?id=2053416>. (12 November 2014).
- Kusumadewi, S. 2004. *Membangaun Jaringan Saraf Tiruan (Menggunakan MATLAB & Excel Link)*. Graha Ilmu: Yogyakarta.
- Prasetyo, E. 2012. *Konsep dan Aplikasi Menggunakan MATLAB*. Penerbit Andi: Yogyakarta.
- Siang, J. J. 2009. *Jaringan Syaraf Tiruan dan Pemogramanya menggunakan MATLAB (Ed. II)*. Yogyakarta: Andi Offset.
- Warsito, B. 2009. *Kapita Selekta Statistika Neural Network*. BP Undip: Semarang.