

CLUSTERING DATA PENCEMARAN UDARA SEKTOR INDUSTRI DI JAWA TENGAH DENGAN KOHONEN NEURAL NETWORK

Budi Warsito, Dwi Ispriyanti dan Henny Widayanti

Program Studi Statistika FMIPA UNDIP Semarang

ABSTRACT

Industrial clustering in Central Java based on polutan yielded to be intended in order to obtaine an industrial group as information in development wisdom specially at Central Java Province. The method that is selected in industrial clustering is Kohonen Artificial Neural Network. An Artificial Neural Network is configured for a specific application, such as pattern recognition or data classification, through a learning process. Kohonen Neural Network can be used in data clustering through unsupervised learning. This network will divide the input pattern into some cluster, based on trained weight. Then this weight will be updated until it can classified itself into the class needed. This paper will present the result of the air contamination data clustering at industrial sector in Central Java at the year 2006 using Kohonen Neural Network. The result of this clustering is industrial clustering, based on polutan yielded, become three clusters.

Key Words : Kohonen, Cluster, air pollution

1. PENDAHULUAN

Pembangunan di Indonesia khususnya pada sektor industri telah membawa dampak bagi kehidupan manusia, baik dampak positif maupun negatif. Salah satu dampak negatif yang ditimbulkan adalah pencemaran udara yang dirasakan semakin meningkat. Dengan meningkatnya beban pencemaran udara sebagai efek negatif dari kegiatan industri, diperlukan pengelompokan industri berdasarkan beban polutan, sehingga dapat diketahui hasilnya berupa kelompok-kelompok industri sebagai informasi dalam kebijaksanaan pembangunan khususnya pada Propinsi Jawa Tengah.

Pengelompokan industri berdasarkan beban polutan dapat dilakukan dengan menggunakan analisis cluster dengan memanfaatkan Jaringan Syaraf Tiruan (*Artificial Neural Network*). Jaringan Syaraf Tiruan (*Artificial Neural Network*) sudah mulai banyak dimanfaatkan sebagai solusi terhadap berbagai macam kasus yang muncul beberapa dekade terakhir. Sejarah ANN menunjukkan pembahasan terhadap masalah ini muncul sekitar tahun 1990-an namun implementasinya baru banyak muncul beberapa dekade terakhir.

Jaringan Syaraf Tiruan (JST) Kohonen merupakan suatu jaringan yang mempunyai kemampuan memetakan pola masukan melalui model pembelajaran tanpa

pengawasan (*unsupervised learning*). Pemetaan yang dihasilkan akan menunjukkan hubungan keterkaitan di antara pola-pola masukan tersebut dalam suatu representasi yang lebih ringkas dari data aslinya dengan mempertahankan hubungan topologinya. (Stergiou, L. 1989)

Menurut Siang (1990), pemanfaatan Jaringan Kohonen dalam pembuatan peta masukan dilakukan dengan membagi pola masukan ke dalam beberapa kelompok (cluster). Dalam kasus pencemaran udara di Jawa Tengah Jaringan Kohonen dapat digunakan untuk mengetahui pengelompokan industri berdasarkan beban polutan yang dihasilkan.

2. JARINGAN SYARAF TIRUAN

Jaringan Syaraf Tiruan (*Artificial Neural Network*) adalah suatu paradigma pengolahan informasi yang diilhami oleh sistem biologi yaitu neuron, seperti otak yang memproses informasi. Kunci Jaringan Syaraf Tiruan adalah struktur sistem pengolahan informasi, yang terdiri atas sejumlah unsur-unsur (syaraf) yang bekerja saling berhubungan untuk memecahkan permasalahan spesifik.

Jaringan Syaraf Tiruan dibangun untuk meniru cara kerja otak manusia. Seperti halnya otak manusia yang terdiri dari sekumpulan sel syaraf (*neuron*), jaringan syaraf juga terdiri dari beberapa *neuron* dan

terdapat hubungan antara *neuron-neuron* tersebut. *Neuron-neuron* tersebut akan memindahkan informasi yang diterima melalui sambungan keluarnya menuju *neuron-neuron* yang lain. Pada jaringan syaraf, hubungan ini dikenal dengan nama bobot. Informasi tersebut disimpan pada nilai tertentu pada bobot.

Proses pembelajaran terhadap perubahan bobot dalam Jaringan Syaraf Tiruan ada dua, yaitu :

(i) Pembelajaran terawasi (*supervised learning*)

Metode pembelajaran pada jaringan syaraf disebut terawasi jika output yang diharapkan telah diketahui sebelumnya. Pada proses pembelajaran, satu pola input akan diberikan ke satu *neuron* pada lapisan input. Pola ini akan dirambatkan di sepanjang jaringan syaraf hingga sampai ke *neuron* pada lapisan output. Lapisan output ini akan membangkitkan pola output yang nantinya akan dicocokkan dengan pola output targetnya. Apabila terjadi perbedaan antara pola output hasil pembelajaran dengan pola target, maka akan muncul error. Apabila nilai error cukup besar, mengindikasikan bahwa masih perlu dilakukan lebih banyak pembelajaran lagi.

(ii) Pembelajaran tak terawasi (*unsupervised learning*)

Pada metode pembelajaran yang tak terawasi ini tidak memerlukan target output. Pada metode ini, tidak dapat ditentukan hasil yang seperti apakah yang diharapkan selama proses pembelajaran. Selama proses pembelajaran, nilai bobot disusun dalam suatu range tertentu tergantung pada nilai input yang diberikan. Tujuan pembelajaran ini adalah mengelompokkan unit-unit yang hampir sama dalam suatu area tertentu.

3. JARINGAN KOHONEN

Jaringan Kohonen termasuk dalam pembelajaran tak terawasi (*unsupervised learning*). Jaringan ini pertama kali diperkenalkan oleh Teuvo Kohonen pada tahun 1981. Pada jaringan ini, suatu lapisan yang berisi *neuron-neuron* akan menyusun dirinya sendiri berdasarkan input nilai tertentu dalam suatu kelompok yang

dikenal dengan istilah cluster. Selama proses penyusunan diri, cluster yang memiliki vektor bobot paling cocok dengan pola input (memiliki jarak paling dekat) akan terpilih sebagai pemenang. *Neuron* yang menjadi pemenang beserta *neuron-neuron* tetangganya akan memperbaiki bobot-bobotnya.

Terdapat m unit kelompok yang tersusun dalam arsitektur sinyal-sinyal masukan (input) sejumlah n . Vektor bobot untuk suatu unit kelompok disediakan dari pola-pola masukan yang tergabung dengan kelompok tersebut. Selama proses pengorganisasian sendiri, unit kelompok yang memiliki vektor bobot paling cocok dengan pola masukan (ditandai dengan jarak *Euclidean* paling minimum) dipilih sebagai pemenang. Unit pemenang dan unit tetangganya diperbaharui bobotnya. Setiap *neuron* terkoneksi dengan *neuron* lain yang dihubungkan dengan bobot atau *weight*. Bobot tersebut berisi informasi yang akan digunakan untuk tujuan tertentu.

Algoritma pembelajaran tanpa supervisi pada Jaringan Kohonen untuk diterapkan dalam pengelompokan data (clustering data) adalah sebagai berikut :

1. Tentukan :
 - a. Jumlah Variabel = m
 - b. Jumlah Data = n
 - c. Jumlah Cluster = K
2. Inisialisasi :
 - a. Bobot Input (w_{ij}) :

$$w_{ij} = \frac{\text{Min}P_j + \text{Max}P_j}{2}$$

dengan

w_{ij} = bobot antara variabel input ke- j dengan neuron pada kelas ke- i .

$\text{Min}P_i$ = nilai minimum pada variabel input ke- i .

$\text{Max}P_i$ = nilai maksimum dari variabel input ke- i .

b. Bobot Bias (b_i)

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

dengan

b_i = bobot bias neuron ke- i .

K = jumlah neuron target.

c. Set parameter *learning rate* (α).

d. Set maksimum epoh (MaxEpoh).

3. Set Epoch = 0.
4. Kerjakan jika Epoch < MaxEpoch
 - a. Epoch = Epoch + 1
 - b. Pilih data secara acak, misalnya data terpilih data ke-z.
 - c. Cari jarak antara data ke-z dengan tiap bobot input ke-i (D_i) :

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

Penjumlahan negatif jarak plus bobot bias (a_i) :

$$a_i = - D_i + b_i$$

Mencari a_i terbesar :

- i. $MaxA = \max(a_i)$, dengan $i = 1, 2, \dots, K$.
 - ii. $Idx = 1$, sedemikian hingga $a_i = MaxA$.
- d. Set output neuron ke-i (y_i) :
- $y(i) = 1$; jika $i = idx$.
 $y(i) = 0$; jika $i \neq idx$.

- e. Update bobot yang menuju ke neuron idx :

$$w_{ji}^{baru} = w_{ji}^{lama} + \alpha(x_i - w_{ji}^{lama})$$

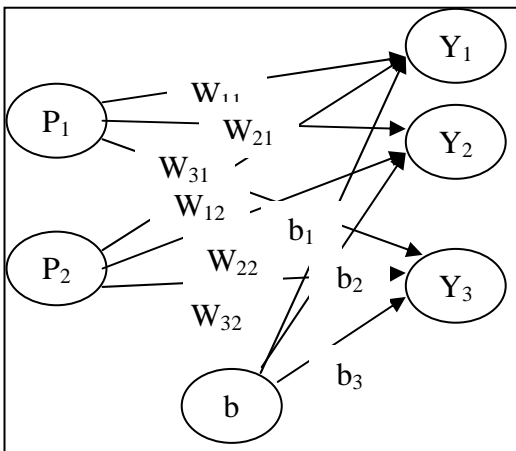
$$w(idx, j) = w(idx, j) + \alpha(p(z, j) - w(idx, j))$$

- f. Update bobot bias :

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

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

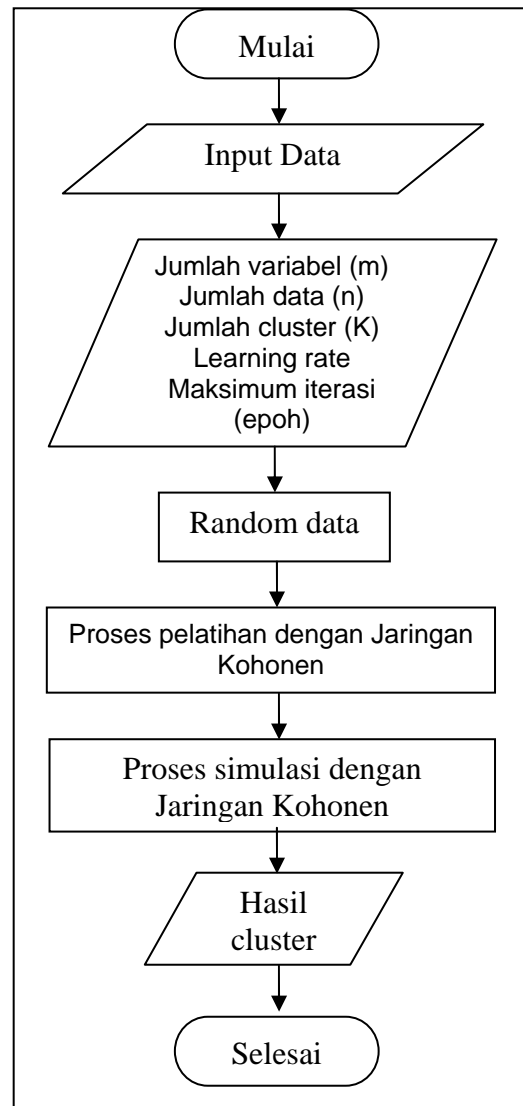
Proses pembelajaran akan berlangsung terus hingga mencapai maksimum epoch.



Gambar 1. Arsitektur Jaringan Kohonen

Jaringan Kohonen dapat mengenali dan mengklasifikasikan pola-pola dengan melakukan pelatihan (*training*) dari pola-pola vektor input (masukan) data dengan vektor bobot sebagai penghubung antara layar masukan dan layar kompetisi dalam proses pelatihan. Dari proses pelatihan jaringan tersebut akan terbentuk cluster-cluster dari pola-pola yang dilatihkan.

Klasifikasi pola-pola tersebut nantinya dapat digunakan sebagai proses pengenalan pola-pola yang diujikan. Proses klasifikasi mencakup cara pengelompokan pola berdasarkan keserupaan ciri yang dimilikinya (*clustering*) dan pemberian label kelas atas masing-masing kelompok tersebut. Diagram alir implementasi pada Jaringan Kohonen disajikan pada gambar berikut.



Gambar 2. Diagram Implementasi Jaringan Kohonen

4. TERAPAN PADA DATA PENCEMARAN UDARA

Learning rate (α) : 0.5

Berikut merupakan informasi awal yang digunakan dalam clustering data industri sumber pencemaran udara di Jawa Tengah tahun 2006 menggunakan jaringan syaraf tiruan Kohonen Neural Network :

Data beban pencemaran udara pada sektor industri di Jawa Tengah tahun 2006 (ton/tahun) secara lengkap disajikan pada tabel 1. Pengambilan data dilakukan oleh Badan Pengelolaan dan Pengendalian Dampak Lingkungan (Bappedal) Propinsi Jawa Tengah Jalan Setiabudi Sronдол Komplek Diklat Propinsi Jawa Tengah.

Jumlah data (n) : 14
 Jumlah variabel input (m) : 5
 Jumlah cluster diinginkan (K) : 3

Tabel 1. Data beban pencemaran udara sektor industri di Jawa Tengah tahun 2006 (ton/tahun)

No	INDUSTRI	DEBU	SO ₂	NO ₂	HC	CO
1	Makanan	17.212,900	160.083,878	74.816,737	3.219,487	399,424
2	Minuman	14.783,549	138.034,875	63.495,986	2.775,775	344,361
3	Tekstil	94.220,918	876.275,960	409.535,961	17.621,236	2.186,081
4	Kayu	37.746,008	350.356,787	163.742,598	19.178,572	874,048
5	Olahan kayu	64.938,164	585.197,249	273.497,539	11.767,867	1.459,915
6	Kertas	4.995,420	490.970,353	21.712,842	934,244	133,504
7	Kimia dasar	6.127,549	56.987,602	27.825,416	1.145,976	142,169
8	Non logam	5.247,509	49.386,337	23.081,177	993,120	123,205
9	Semen	518,181	4.816,351	2.250,968	96,852	12,015
10	Kapur dan gips	10.085,093	19.659,868	10.865,165	1.889,810	234,446
11	Logam dasar	4.012,867	36.194,179	22.328,706	1.018,972	90,295
12	Hasil-hasil olahan logam	7.407,925	68.895,388	32.198,916	1.385,432	171,037
13	Rumah sakit	64,523	59,875	288,565	12,039	1,492
14	Perhotelan	99,836	92,747	433,466	18,650	2,312
Jumlah Total		267.460,442	2.837.011,449	1.126.074,041	62.058,032	6.174,306

Sumber : BAPPEDAL Propinsi Jawa Tengah

Berdasarkan input data pada tabel 1 kemudian dilakukan pelatihan jaringan masing-masing dengan 1.000 dan 10.000 epoch menggunakan paket program Matlab 7.0. Hasil pelatihan jaringan dari masing-masing epoch diperoleh keluaran sebagai berikut :

Tabel 3. Hasil pembacaan clustering data untuk 10.000 epoch :

Tabel 2. Hasil pembacaan clustering data untuk 1.000 epoch :

Cluster ke-	Masukan ke-
1	1, 2, 7, 8, 9, 10, 11, 12, 13, 14
2	3
3	4, 5, 6

Cluster ke-	Masukan ke-
1	1, 2, 7, 8, 9, 10, 11, 12, 13, 14
2	4, 5, 6
3	3

Dari tabel 2 dan tabel 3 dapat dilihat bahwa hasil pelatihan dengan 1.000 dan 10.000 epoch menghasilkan cluster dengan anggota yang konsisten. Perbedaan hanya terletak pada penempatan cluster yang disebabkan oleh faktor pemilihan data secara acak (randomisasi). Secara detail hasil clustering disajikan pada tabel 4 berikut.

Tabel 4. Hasil clustering data pencemaran udara sektor industri di Jawa Tengah tahun 2006 :

Cluster ke-	Jenis Industri
1	Makanan, Minuman, Kimia Dasar, Non Logam, Semen, Kapur dan Gips, Logam Dasar, Hasil-hasil Olahan Logam, Rumah Sakit, Perhotelan
2	Tekstil
3	Kayu, Olahan Kayu, Kertas

5. KESIMPULAN

Berdasarkan pembahasan

1. Penyelesaian permasalahan clustering data menggunakan Jaringan Kohonen dipengaruhi oleh parameter-parameter pelatihan seperti jumlah cluster yang akan dibentuk, learning rate, maksimum iterasi (epoh) sehingga jika proses pelatihan dilakukan beberapa kali dengan data masukan yang sama, akan berpengaruh pada clustering data yang dihasilkan. Hal ini disebabkan pada Algoritma Kohonen untuk clustering data terdapat pemilihan data secara acak (randomisasi).
2. Hasil clustering pada data beban pencemaran udara pada sektor industri di Jawa Tengah pada tahun 2006 dengan menggunakan Jaringan Kohonen diperoleh :
 - a. Cluster I : industri tekstil.
 - b. Cluster II : industri makanan, industri minuman, industri kimia dasar, industri non logam, industri semen, industri kapur dan gips, industri logam dasar, industri hasil-hasil olahan logam, industri rumah sakit, dan industri perhotelan.
 - c. Cluster III : industri kayu, industri olahan kayu, dan industri kertas.

UCAPAN TERIMA KASIH

Pada kesempatan ini penulis menyampaikan ucapan terima kasih kepada Pengelola Program Hibah Penelitian PHK A2 Jurusan Matematika FMIPA UNDIP yang telah membiayai penelitian yang dilakukan penulis sehingga tulisan ini dapat diterbitkan.

DAFTAR PUSTAKA

- Bappedal. 2006. *Beban Pencemaran Udara Menurut Sumber*. Bappedal : Jawa Tengah
- Fausett, L. 1994. *Fundamentals of Neural Networks, Architectures, Algorithms, and Applications*. Prentice Hall : New Jersey
- Haryanto, L. 2004. *Membangun Perangkat Lunak Untuk Data Clustering Menggunakan Jaringan Syaraf Algoritma Kohonen Self Organizing*, Jurusan Teknik Informatika FT UII : Yogyakarta
- <http://www.cse.cuhk.edu.hk/~idel98/tutorial.html>
- Germano, T. 1999. *Self Organizing Maps Tutor*. Springer Verlag : New York (diakses terakhir tanggal 11 April 2007 09.14 WIB)
- <http://www.heatonresearch.com/articles/6/page2.html>
- Heaton, J. 2003. *Introduction to Neural Network with Java*. (diakses terakhir tanggal 30 Maret 2007 12.27 WIB)
- <http://www.emsl.pnl.gov:2080/docs/cie/neural/neural.homepage.html>
- Neural Networks at Pacific Northwest National Laboratory. *Artificial Neural Networks*. (diakses terakhir tanggal 30 April 2007 16.52 WIB)
- http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html
- Stergiou, C. and Siganos, D. 1989. *Neural Networks*. Pacific Northwest National Laboratory : New York (diakses terakhir tanggal 30 Maret 2007 12.04 WIB)
- <http://www.cs.stir.ac.uk/~lss/NNIntro.html>
- Smith, L. 1996. *An Introduction to Neural Networks*. Department of Computing and Mathematics University of Stirling : Scotland
- (diakses terakhir tanggal 11 April 2007 10.32 WIB)
- Johnson, R. A and Wichern, D.W. 1982. *Applied Multivariate Statistical Analysis*. Prentice-Hall Inc : New Jersey

Kusumadewi, S. 2003. *Artificial Intelligence Teknik dan Aplikasinya*. Graha Ilmu : Yogyakarta

Sastrawijaya, A. T. 2000. *Pencemaran Lingkungan*. Surabaya : Rineka Cipta

Siang, J. J. 2005. *Jaringan Syaraf Tiruan dan Pemrogramannya Menggunakan Matlab*. Penerbit Andi : Yogyakarta

Sichah, I. A. dan Safitri, D. 2005. *Statistika Multivariat*. Jurusan Matematika Undip : Semarang

Siong, A. W. dan Resmana. 1999. *Pengenalan Citra Objek Sederhana dengan Menggunakan Metode Jaringan Saraf Tiruan SOM*, jurnal Pusat Komputer Jurusan Teknik Elektro Universitas Kristen Petra : Surabaya