

Penerapan *Data Science* Menggunakan *Artificial Neural Network* (ANN) Metode *Self Organizing Mapping* (SOM) untuk Klasifikasi Industri

Ismail¹

¹)Program Studi Analisis Kimia, Politeknik AKA Bogor
Jl. Pangeran Sogiri No.283, Tanah Baru, Bogor Utara, Kota Bogor, Jawa Barat 16154

^{*}Email: ismail.abusyafiq@gmail.com

(Received : 5 Desember 2021; Accepted: 12 Desember 2021; Published: 21 Desember 2021)

Abstrak

Artificial Neural Network (ANN) sebagai sebuah metode analitik dalam *data science* memiliki kemampuan pada bidang klasifikasi, asosiasi, *self organizing* dan optimasi. Dengan kemampuan tersebut ANN dapat menciptakan suatu pola pengetahuan melalui pengaturan diri atau kemampuan belajar (*self-organizing*) bahkan mampu memetakan kumpulan data berdasarkan ciri-ciri kesamaan yang ada pada data tersebut dengan mekanisme *Self Organizing Mapping* (SOM). Kemampuan SOM dalam melakukan pemetaan terhitung cukup valid dan rendah bias, karena mengkalkulasi secara mandiri dan cepat berdasarkan kesamaan ciri atau karakter data yang diinput. Pada penelitian ini SOM diterapkan untuk pemetaan industri pada industri mesin pertanian dan kehutanan. Mengingat selama ini kekuatan industri mesin pertanian dan kehutanan dengan kode Kelompok Lapangan Usaha Industri (KLUI) 2921 belum dipetakan dengan sempurna. Data industri yang menjadi input pada program terlebih dahulu ditransformasi dengan aktivasi sigmoid biner dan selanjutnya dimasukkan ke dalam program jaringan kompetitive kohonen. Berdasarkan perhitungan metode SOM dihasilkan klasifikasi industri dengan jenis menengah dan besar dengan capaian *epoch* 1000.

Kata kunci: Data Science; Artificial Neural Network (ANN); Self Organizing Mapping (SOM); Klasifikasi Industri.

Abstract

Artificial Neural Network (ANN), as an analytical method in *data science*, has capabilities in classification, association, *self-organizing* and optimization. With this capability, the ANN can create a pattern of knowledge through self-regulation or self-organizing ability and map data sets based on the similarity characteristics that exist in the data with the *Self Organizing Mapping* (SOM) mechanism. SOM's ability to perform mapping is considered valid and low bias because it calculates independently and quickly based on the similarity of characteristics or characters of the input data. This study applied SOM for industrial mapping in the agricultural and forestry machinery industry. Given that so far, the strength of the agricultural and forestry machinery industry with the code Industrial Business Field Group (KLUI) 2921 has not been mapped perfectly. The industrial data input to the program is first transformed by binary sigmoid activation and then entered into a Kohonen competitive network program. Based on the calculation of the SOM method, industrial classifications are produced with medium and large types with 1000 epoch achievements.

Keywords: Data Science; Artificial Neural Network (ANN); Self Organizing Mapping (SOM); Industry classification

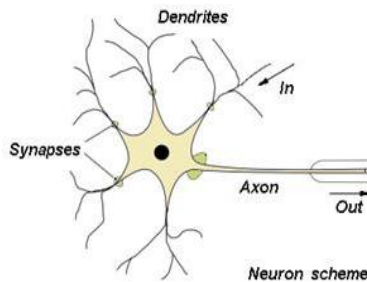
PENDAHULUAN

Warren McCulloch dan Walter Pitts telah memperkenalkan perhitungan model *Artificial Neural Network* (ANN) pada pertama kalinya pada tahun 1943. Ilmu tentang ANN yang merupakan kategori ilmu *data science* terus mengalami

perkembangan. Secara umum ANN mampu mengadopsi kecerdasan dari kemampuan otak manusia yang dapat memberikan stimulasi, rangsangan, pemrosesan stimulus, dan pemberian output. Output diperoleh dari variasi stimulasi dan proses yang terjadi di dalam otak manusia.

Kemampuan manusia dalam memproses informasi merupakan hasil kompleksitas proses di dalam otak. Pola kerja jaringan syaraf pada otak manusia terdiri atas 3 (tiga) organ utama seperti pada Gambar 1 yaitu:

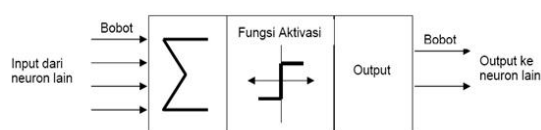
1. dendrit (*dendrites*) berfungsi untuk mengirimkan impuls yang diterima ke badan sel syaraf;
2. akson (*axon*) yang berfungsi untuk mengirimkan impuls dari badan sel ke jaringan lain; dan
3. sinapsis yang berfungsi sebagai unit fungsional di antara dua sel syaraf. (Bishop, 2007)



Gambar 1. Struktur Neuron pada Otak

Menurut Huang et al (2011) berdasarkan karakteristik pola kerja jaringan syaraf maka pembentukan karakteristik dari ANN disesuaikan berdasarkan pola hubungan antar neuron, metode penentuan bobot dari tiap koneksi, dan fungsi aktivasinya seperti pada Gambar 2. Pada Gambar 2 tersebut menjelaskan struktur ANN secara mendasar, meskipun pada kenyataannya dapat lebih kompleks. Struktur kerja ANN sedikitnya memuat:

1. Input, berfungsi seperti dendrite
2. Output, berfungsi seperti akson
3. Fungsi aktivasi, berfungsi seperti sinapsis



Gambar 2 Struktur ANN

Algoritma jaringan di ANN terdiri atas *supervised learning* dan *unsupervised learning*. Ciri algoritma *supervised learning* diantaranya adalah memerlukan target output yang diharapkan untuk diketahui sebelumnya dalam proses pembelajaran, serta setiap pola input beserta target output yang ditentukan dijadikan pasangan pembelajaran. Algoritma *neural network* yang termasuk dalam *supervised learning* antara lain *hebb*, *perceptron* dan *back propagation*. Sedangkan algoritma *unsupervised learning* tidak memerlukan target output dan tidak dapat ditentukan hasilnya seperti apa yang diharapkan selama proses pembelajaran. Tujuan pembelajaran *unsupervised learning* adalah mengelompokkan unit-unit yang hampir sama

dalam suatu area tertentu. ANN termasuk dalam *unsupervised learning* antara lain: *kohonen self-organizing maps*, *learning vector quantization*, dan *counter-propagation* (Siang, 2005).

Fungsi Aktivasi di ANN dapat berupa fungsi tangga biner yang mengkonversikan input dari suatu variabel yang bernilai kontinu ke suatu output biner (0 atau 1); atau dapat berupa fungsi linear (identitas) yang memiliki nilai output sama dengan nilai inputnya. Fungsi linear dirumuskan sebagai, $y = x$; atau fungsi sigmoid biner yang mempunyai nilai pada *range* 0 sampai 1; ataupun fungsi sigmoid bipolar yang mempunyai nilai pada *range* antara -1 sampai 1. Berdasarkan studi, ANN memiliki kelebihan dan kekurangan sebagai sebuah metode dalam *data science*. Diantara kelebihannya adalah:

- mampu mengakuisisi pengetahuan walau tidak ada kepastian;
- mampu melakukan generalisasi dan ekstraksi dari suatu pola data tertentu;
- ANN dapat menciptakan suatu pola pengetahuan melalui pengaturan diri atau kemampuan belajar (*self organizing*);
- memiliki *fault tolerance*, gangguan dapat dianggap sebagai *noise* saja;
- kemampuan perhitungan secara paralel sehingga proses lebih singkat

Sedangkan kekurangan ANN adalah:

- proses kerja mengikuti prinsip *black box*;
- kurang mampu melakukan operasi-operasi numerik dengan presisi tinggi;
- kurang mampu melakukan operasi algoritma aritmatik, operasi logika dan simbolis;
- proses *training* yang mungkin terjadi dalam waktu yang sangat lama untuk jumlah data yang besar .

ANN sebagai sebuah metode analitik dalam *data science* memiliki kemampuan pada bidang:

- klasifikasi: memilih suatu input data ke dalam kategori tertentu yang sudah ditetapkan;
- asosiasi: menggambarkan suatu obyek secara keseluruhan hanya dengan bagian dari obyek lain;
- *self organizing*: kemampuan mengolah data-data input tanpa harus mempunyai target; dan
- optimasi: menemukan suatu jawaban terbaik sehingga mampu meminimalisasi fungsi biaya.

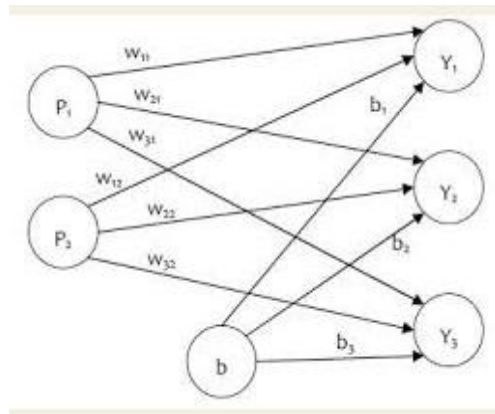
Berdasarkan kajian tentang ANN, dengan segala kelebihannya sebagai metode algoritma dalam *data science*, maka dapat diuji cobakan aplikasinya dalam bidang pemetaan industri pada industri mesin pertanian dan kehutanan. Mengingat selama ini kekuatan industri mesin pertanian dan kehutanan belum dipetakan dengan sempurna. Industri mesin pertanian dan kehutanan merupakan bagian dari industri permesinan dan peralatan dengan kode Kelompok Lapangan Usaha Industri (KLUI) 2921.

Pemetaan ini berguna untuk penentuan kekuatan industri mesin lebih lanjut berkaitan dengan daya saing industri mesin Indonesia. Dengan kondisi tersebut maka perlu dilakukan usaha pemetaan industri khususnya perusahaan penghasil mesin pertanian dan kehutanan dengan mengklasifikasikannya menurut skala industri. Skala industri yang digunakan adalah besar, menengah dan kecil. Data yang menjadi acuan adalah berasal dari Badan Pusat Statistik (BPS).

METODE PENELITIAN

Metode penelitian ini disusun dengan urutan langkah sebagai berikut:

1. Mengumpulkan data industri berdasarkan nilai produksi, nilai tambah dan jumlah tenaga kerja yang dimiliki oleh industri dengan KLUI 2921 dan menjadikannya sebagai data input;
2. Mentransformasi data input dengan fungsi aktivasi sigmoid biner
3. Menentukan tipe network, yakni kohonen atau jaringan kompetitif berdasarkan metode clustering dengan *Neural Network Unsupervised Learning* tipe *Self Organizing Mapping* (SOM);
4. Topologi SOM dibentuk mengikuti pola pada Gambar 3.



Keterangan:

P_1, P_2 = input. $w_{11} \dots w_{32}$ = bobot. $b_1 \dots b_3$ = bias. $Y_1 \dots Y_3$ = Output

Gambar 3. Topologi SOM

5. *Self-Organizing Map* ini termasuk ke dalam metode pelatihan tak terbimbing. Dalam metode belajar tak terbimbing, jaringan syaraf tidak memiliki suatu target keluaran tertentu. Jaringan syaraf akan mengelompokkan vektor masukan bersama tanpa menggunakan data latihan untuk mencirikan ke kelompok masukan. Sehingga yang disediakan hanyalah suatu deret vektor-vektor masukan, tanpa adanya vektor-vektor keluaran target.
6. Algoritma SOM disusun sebagai berikut:
 - a) Inisialisasi vektor bobot, w ;
 - b) Menetapkan topologi parameter tetangga;
 - c) Menetapkan parameter kecepatan pembelajaran (*learning rate*);

- d) Menentukan bobot vektor yang paling dekat dengan vektor input. Hal ini dilakukan dengan mencari bobot vektor yang memiliki jarak Euclidean terdekat (*Euclidean distance*);
- e) Menghitung bobot vektor yang telah disesuaikan (*adapt the weights*); dan
- f) Kecepatan pembelajaran disesuaikan.

HASIL DAN PEMBAHASAN

Hasil pengumpulan data industri dengan kode Kelompok Lapangan Usaha Industri (KLUI) 2921 disajikan dalam format tabel kapasitas dan volume industri. Kapasitas industri disusun berdasarkan: nilai produksi, nilai tambah, dan jumlah tenaga kerja (naker) yang dimiliki oleh industri tersebut sebagaimana ditunjukkan pada pada Tabel 1.

Tabel 1 Kapasitas Industri KLUI 2921

KLUI	Nilai Produksi		Jumlah Naker
		Tambah	
2921	276000	158922	22
2921	384000	337107	21
2921	3087926	1615115	81
2921	0	1707291	37
2921	91235562	14972298	377
2921	179040	120390	20
2921	103889373	19979320	40
2921	13639888	1179138	34
2921	503710	99892	26
2921	425860	196454	24
2921	76448827	12653023	429
2921	169557214	78950352	368
2921	807632	527976	28
2921	2234800	910702	29
2921	81000	169340	36
2921	412500	180220	20
2921	281328	148148	21
2921	720000	193572	30
2921	1346870	485119	57
2921	348500	135424	21
2921	226250	84242	28

Selanjutnya data ditransformasi dengan keluaran data mengikuti pola sigmoid biner dari rentang 0 sampai 1. Perhitungannya mengikuti rumus sebagai berikut:

$$data\ transformasi = \frac{0.8 \times (data - data\ min)}{data\ max - data\ min} + 0.1$$

Data transformasi disajikan sebagaimana ditunjukkan pada Tabel 2 berikut:

Tabel 2 Data Hasil Transformasi

Nilai produksi	Nilai tambah	Jumlah tenaga kerja
0.101302215	0.100757537	0.10391198
0.101811778	0.102565005	0.10195599
0.114569364	0.11552883	0.219315403
0.1	0.116463842	0.133251834
0.53046502	0.251021076	0.798288509
0.100844741	0.100366677	0.1
0.59016787	0.301811176	0.139119804
0.16435533	0.111106378	0.127383863
0.10237659	0.10015875	0.111735941
0.10200928	0.101138253	0.107823961
0.46069867	0.227494875	0.9
0.9	0.9	0.780684597
0.103810546	0.104501137	0.115647922
0.110544169	0.108383424	0.117603912
0.100382172	0.100863215	0.131295844
0.101946246	0.100973579	0.1
0.101327354	0.100648248	0.10195599
0.103397083	0.101109019	0.119559902
0.106354764	0.104066406	0.172371638
0.101644283	0.100519179	0.10195599
0.101067486	0.1	0.115647922

Setelah data ditranformasi selajutnya menentukan fungsi transfer. Fungsi transfer yang digunakan adalah *Competitive*. Jaringan ditetapkan melakukan iterasi belajar sebanyak 1000 *epoch*. *Syntax* pemrograman disusun sebagai berikut:

- ▲ net=newc([0 1; 0 1; 0 1], 3);
- ▲ net.IW {1,1,1}
- ▲ net.b{1}
- ▲ net=train(net,p);
- ▲ b=sim(net,p)
- ▲ ac=vec2ind(b)
- ▲ net=init(net);
- ▲ net.trainParam.epochs=1000;
- ▲ net=train(net,p);
- ▲ b=sim(net,p)
- ▲ ac=vec2ind(b)

Selanjutnya akan diperoleh output program sebagai berikut:

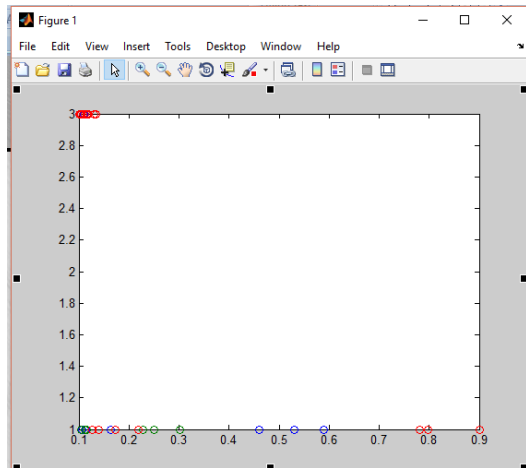
- ▲ net=init(net);
- ▲ net.trainParam.epochs=1000;
- ▲ net=train(net,p);
- ▲ TRAINR, Epoch 0/1000TRAINR, Epoch
- 25/1000TRAINR, Epoch
- 50/1000TRAINR, Epoch
- 75/1000TRAINR, Epoch
- 100/1000TRAINR, Epoch
- 125/1000TRAINR, Epoch
- 150/000TRAINR, Epoch
- 175/1000TRAINR, Epoch
- 200/1000TRAINR, Epoch
- 225/1000TRAINR, Epoch
- 250/1000TRAINR, Epoch
- 275/1000TRAINR, Epoch

300/1000TRAINR, Epoch
 325/1000TRAINR, Epoch
 350/1000TRAINR, Epoch
 375/1000TRAINR, Epoch
 400/1000TRAINR, Epoch
 425/1000TRAINR, Epoch
 450/1000TRAINR, Epoch
 475/1000TRAINR, Epoch
 500/1000TRAINR, Epoch
 525/1000TRAINR, Epoch
 550/1000TRAINR, Epoch
 575/1000TRAINR, Epoch
 600/1000TRAINR, Epoch
 625/1000TRAINR, Epoch
 650/1000TRAINR, Epoch
 675/1000TRAINR, Epoch
 700/1000TRAINR, Epoch
 725/1000TRAINR, Epoch
 750/1000TRAINR, Epoch
 775/1000TRAINR, Epoch
 800/1000TRAINR, Epoch
 825/1000TRAINR, Epoch
 850/1000TRAINR, Epoch
 875/1000TRAINR, Epoch
 900/1000TRAINR, Epoch
 925/1000TRAINR, Epoch
 950/1000TRAINR, Epoch
 975/1000TRAINR, Epoch 1000/1000

- ▲ TRAINR, Maximum epoch reached.
- ▲ b=sim(net,p)
- ▲ b = (3,1) 1 (3,2) 1 (1,3) 1 (3,4)
 1 (1,5) 1 (3,6) 1 (1,7) 1
 (1,8) 1 (3,9) 1 (3,10) 1
 (1,11) 1 (1,12) 1 (3,13) 1
 (3,14) 1 (3,15) 1 (3,16) 1
 (3,17) 1 (3,18) 1 (1,19) 1
 (3,20) 1 (3,21) 1
- ▲ ac=vec2ind(b)
- ▲ ac = 3 3 1 3 1 3 1 1 3 3
 1 1 3 3 3 3 3 3 1 3 3

Hasil klasifikasi secara bagan dua dimensi ditunjukkan pada Gambar 4.

Berdasarkan hasil output program dapat dipetakan klasifikasi industri dengan KLUI 2921 yang terdiri atas 2 (dua) klasifikasi industri dengan jenis/penggolan industri menengah dan besar berdsarkan nilai produksi, nilai tambah, dan jumlah tenaga kerjanya. Hasil pemetaan dan klasifikasi dapat dilihat sebagaimana ditunjukkan pada Tabel 3.



Gambar 4. Bagan Hasil Klasifikasi

Tabel 3. Hasil Klasifikasi berdasarkan SOM

KLUI	NP	NT	JT	JK	SI
2921	276000	158922	22	3	M
2921	384000	337107	21	3	M
2921	3087926	1615115	81	1	B
2921	0	1707291	37	3	M
2921	91235562	14972298	377	1	B
2921	179040	120390	20	3	M
2921	103889373	19979320	40	1	B
2921	13639888	1179138	34	1	B
2921	503710	99892	26	3	M
2921	425860	196454	24	3	M
2921	76448827	12653023	429	1	B
2921	169557214	78950352	368	1	B
2921	807632	527976	28	3	M
2921	2234800	910702	29	3	M
2921	81000	169340	36	3	M
2921	412500	180220	20	3	M
2921	281328	148148	21	3	M
2921	720000	193572	30	3	M
2921	1346870	485119	57	1	B
2921	348500	135424	21	3	M
2921	226250	84242	28	3	M

Ket:

NP: Nilai Produksi

NT: Nilai Tambah

JT: Jumlah Tenaga kerja

JK: Jenis Kelompok

SI: Skala Industri

M: Menengah

B: Besar

KESIMPULAN

Berdasarkan hasil perhitungan dan analisis diperoleh hasil 2 (dua) buah kelompok skala industri yaitu 1 dan 3. Melihat pertimbangan bahwa nilai-nilai parameter kelompok 1 mempunyai nilai transformasi tinggi maka kelompok 1 dapat dinyatakan sebagai kelompok industri skala besar. Sedangkan kelompok 3 dapat dinyatakan sebagai kelompok industri yang bukan skala besar.

Dengan pertimbangan hasil pada epoch 100 dan realitas bahwa industri mesin pertanian dan kehutanan beroperasi pada skala besar dan menengah dan cenderung jarang pada skala usaha kecil, maka kelompok 3 dinyatakan sebagai industri skala menengah. Dari hasil analisis juga dapat dinyatakan bahwa industri mesin pertanian dan kehutanan tergolong berskala besar atau menengah dapat diurutkan dari faktor: nilai produksinya, nilai tambah dan jumlah tenaga kerjanya.

DAFTAR PUSTAKA

- Bishop, C. (2007). *Natural networks for pattern recognition* (Repr.). Oxford: Oxford University Press.
- Huang, G.-B., Wang, D. H., & Lan, Y. (2011). *Extreme learning machines: A survey*. *International Journal of Machine Learning and Cybernetics*, 2, 107–122. doi:10.1007/513042-011-0019-y.
- Siang, JJ. (2005). *Jaringan Syaraf Tiruan & Pemrogramannya Menggunakan Matlab*. Ikrar Mandiriabakti Publisher.