

## Implementasi Metode *Artificial Neural Network* (ANN) Algoritma *Backpropagation* untuk Klasifikasi Kualitas Udara di Provinsi DKI Jakarta Tahun 2021

Lutfiah Anindya Putri\*, Suwanda

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

\*Luthfiaanindya1@gmail.com, idris1000358@gmail.com

**Abstract.** Classification is the systematic division of an individual into a group or category according to established rules or standards. There are several classification algorithms that can be used to classify data, including logistic regression, discriminant analysis, and Artificial Neural Network (ANN). This study uses an Artificial Neural Network (ANN) classification algorithm because the data used is not normally distributed. ANN is a classification algorithm that is widely used for character recognition problems and is a strong classifier because of its high calculation rate and accuracy. In this thesis, ANN is implemented on air quality classification data for DKI Jakarta in 2021 with particulate (PM<sub>10</sub>), sulfide (SO<sub>2</sub>), carbon monoxide (CO), Ozone (O<sub>3</sub>), and nitrogen dioxide (NO<sub>2</sub>) variables. This method will be calculated based on the accuracy, precision, and recall values obtained from the confused matrix. The most optimal classification results are obtained from the network architecture, which consists of 5 input layers, 4 hidden layers, and 2 output layers with an epoch value of 5000 and a learning rate of 0.001. The network architecture produces an accuracy value of 94%, a precision of 90%, and a recall of 100%.

**Keywords:** *Artificial Neural Network (ANN), Confusion Matrix, Classification, and Air Quality.*

**Abstrak.** Klasifikasi adalah pembagian sistematis dari sebuah individu ke dalam suatu kelompok atau kategori menurut aturan atau standar yang ditetapkan. Terdapat beberapa algoritma klasifikasi yang dapat digunakan dalam mengklasifikasi data, diantaranya regresi logistik, analisis diskriminan, dan *Artificial Neural Network* (ANN). Penelitian ini menggunakan algoritma klasifikasi *Artificial Neural Network* (ANN) karena data yang digunakan tidak berdistribusi normal. ANN merupakan salah satu algoritma klasifikasi yang banyak digunakan untuk masalah pengenalan karakter dan merupakan *classifier* yang kuat karena tingkat perhitungan dan keakuratannya yang tinggi. Pada skripsi ini, ANN diimplementasikan pada data klasifikasi kualitas udara DKI Jakarta Tahun 2021 dengan variabel partikulat (PM<sub>10</sub>), sulfida (SO<sub>2</sub>), karbon monoksida (CO), Ozon (O<sub>3</sub>), dan nitrogen dioksida (NO<sub>2</sub>). Metode tersebut akan dihitung berdasarkan ukuran nilai *accuracy*, *precision*, dan *recall* yang didapat dari *confusion matrix*. Hasil pengklasifikasian yang paling optimal didapat dari arsitektur jaringan yang terdiri dari 5 *input layer*, 4 *hidden layer*, dan 2 *output layer* dengan nilai *epoch* 5000 dan *learning rate* 0.001. Arsitektur jaringan tersebut menghasilkan nilai *accuracy* sebesar 94%, *precision* sebesar 90%, dan *recall* sebesar 100%.

**Kata Kunci:** *Artificial Neural Network (ANN), Confusion Matrix, Klasifikasi, Kualitas Udara.*

## A. Pendahuluan

Setiap individu memiliki ciri dan karakteristiknya sendiri. Klasifikasi memungkinkan kita untuk mengklasifikasikan individu ke dalam kelompok atau kategori menurut aturan atau norma yang ditetapkan. Menurut Lasa HS (2008), klasifikasi adalah suatu sistem atau aturan dimana bahan pustaka dibagi ke dalam kelas atau kategori tertentu berdasarkan topik tertentu.

*Artificial Neural Network* (ANN) adalah algoritma klasifikasi yang biasa digunakan untuk masalah pengenalan karakter dan merupakan pengklasifikasi yang kuat karena tingkat komputasi dan akurasi yang tinggi (Matondang, 2013). Struktur ANN terdiri dari tiga lapisan, yaitu *input layer*, *hidden layer*, dan *output layer*. Setiap lapisan memiliki beberapa neuron yang terhubung ke lapisan berikutnya (Ripley, 1996). Algoritma *backpropagation* digunakan dalam penelitian ini karena merupakan algoritma *supervised* dimana masukan dan keluaran diketahui terlebih dahulu (Kusumadewi, 2004). Keunggulan dari algoritma ini adalah dapat melatih jaringan untuk mencapai keseimbangan antara kemampuan jaringan untuk mengenali pola yang digunakan selama pelatihan dan merespon pola *input* dengan benar (Effendi, 2013). Oleh karena itu, ANN diharapkan memiliki keunggulan dan efisiensi yang sama atau hampir sama dengan jaringan syaraf manusia dalam mengklasifikasikan informasi (Effendi, 2013) (Muhammad & Aceng, 2023). Beberapa hal yang harus ditentukan dalam algoritma *backpropagation*, yaitu fungsi aktivasi digunakan untuk mengaktifkan dan menonaktifkan neuron, *learning rate* adalah parameter untuk menghitung nilai koreksi bobot selama pelatihan, dan *epoch* merupakan kriteria untuk menghentikan pelatihan (Utami, Aprilia T.W; Ulama, 2015).

ISPU (Indeks Standar Pencemar Udara) adalah angka tanpa satuan yang menjelaskan keadaan kualitas udara di suatu tempat menurut pengaruhnya terhadap kesehatan masyarakat, nilai estetika dan kehidupan orang lain. Tujuan ISPU adalah untuk memberikan informasi kepada masyarakat tentang konsistensi data kualitas udara di suatu tempat pada waktu tertentu, serta bahan operasional yang terkait dengan pengendalian pencemaran udara (Kusnandar, 2020). Penelitian ini bertujuan untuk mengetahui arsitektur jaringan yang paling optimal untuk klasifikasi kualitas udara dan mengukur nilai *accuracy*, *precision*, dan *recall*.

## B. Metodologi Penelitian

Pada penelitian ini langkah pertama adalah melakukan *preprocessing* data dengan mengubah data menjadi numerik, menormalisasi data, dan *split* data menjadi data *training* dan *testing*. Kemudian melakukan langkah-langkah algoritma *backpropagation* dengan terlebih dahulu melakukan inialisasi bobot dan menentukan nilai *learning rate* (0.001, 0.01, dan 0.1), *epoch* (50, 200, 500, 1000, dan 5000), serta jumlah *hidden layer* (1, 2, 3, 4, dan 5). Setelah arsitektur jaringan yang paling optimal didapatkan, kemudian menghitung nilai *accuracy*, *precision*, dan *recall* melalui *confusion matrix*.

## C. Hasil Penelitian dan Pembahasan

### *One-Hot Encoding*

Data awal yang digunakan masih terdapat data kategorik, yaitu pada variabel kategori. Maka dari itu, variabel tersebut akan diubah menjadi data numerik dengan nilai 0 sebagai sedang dan nilai 1 sebagai tidak sehat. Data yang sudah diubah dapat dilihat sebagai berikut:

**Tabel 1.** *One-Hot Encoding*

PM10	SO2	CO	O3	NO2	Kategori
43	58	29	35	65	0
58	86	38	64	80	0
64	93	25	62	86	0
:	:	:	:	:	:
61	54	15	37	29	0
60	53	17	38	44	1
64	52	44	37	53	0

### Normalisasi Data

Tahap selanjutnya adalah melakukan normalisasi dengan menggunakan metode normalisasi *Min-Max*. Contoh normalisasi data ke-1 pada variabel PM<sub>10</sub> dengan rumus sebagai berikut:

$$z = \frac{x - \min(x)}{\max(x) - \min(x)} = \frac{43 - 19}{179 - 19} = 0.15$$

Semua hasil data yang sudah dinormalisasi terdapat pada **Tabel 2**.

**Tabel 2.** Perhitungan Normalisasi

PM10	SO2	CO	O3	NO2
0.15	0.23596	0.55	0.1145	0.448
0.24375	0.55056	0.775	0.33588	0.568
0.28125	0.62921	0.45	0.32061	0.616
0.19375	0.33708	0.425	0.08397	0.544
:	:	:	:	:
0.25	0.17978	0.225	0.10687	0.192
0.2625	0.19101	0.2	0.12977	0.16
0.25625	0.17978	0.25	0.1374	0.28
0.28125	0.16854	0.925	0.12977	0.352

### Split Data

Sebelum melakukan proses analisis data menggunakan metode ANN, data terlebih dahulu dibagi menjadi data training dan testing. Pembagian data menggunakan *k-fold cross validation* dengan nilai  $k = 10$ . Terdapat sebanyak 329 untuk data *training* dan 36 untuk data *testing*.

**Tabel 3.** Data *training*

X					Y
PM10	SO2	CO	O3	NO2	
0.15	0.235	0.55	0.114	0.448	0
0.243	0.55	0.775	0.335	0.568	0
0.193	0.337	0.425	0.083	0.544	0
:	:	:	:	:	:
0.262	0.191	0.2	0.129	0.16	0
0.256	0.179	0.25	0.137	0.28	1
0.281	0.168	0.925	0.129	0.352	0

**Tabel 4.** Data *testing*

X					Y
PM10	SO2	CO	O3	NO2	
0.281	0.1685	0.225	0.068	0.2	1
0.293	0.157	0.525	0.183	0.256	1
0.268	0.089	0.15	0.137	0.072	0
0.306	0.213	0.4	0.152	0.336	1
:	:	:	:	:	:
0.218	0.157	0.275	0.221	0.208	0
0.437	1	1	0.312	0.76	1
0.206	0.258	0.05	0.229	0.152	0
0.225	0.157	0.075	0.351	0.096	0

**Hasil Analisis Artificial Neural Network (ANN)**

Untuk mengklasifikasi kualitas udara menggunakan algoritma *backpropagation* maka harus melatih model terlebih dahulu. Dalam penelitian ini dibuat berbagai model *backpropagation* dengan nilai *learning rate*, jumlah *hidden layer*, dan nilai *epoch* yang berbeda dan akan dievaluasi menggunakan *k-fold cross validation* dengan nilai  $k = 10$ . Jumlah *hidden layer* yang akan dicoba 1 sampai 5, nilai *learning rate* yang diuji adalah 0.01, 0.001, dan 0.1, serta terdapat beberapa nilai *epoch* yang digunakan, yaitu 50, 200, 500, 1000, dan 5000.

**Tabel 5.** Perhitungan hasil dengan *hidden layer* 1 dan 2

		<i>hidden layer</i>					
		1			2		
<i>epoch</i>	<i>LR</i>	<i>accuracy</i>	<i>precision</i>	<i>recall</i>	<i>accuracy</i>	<i>precision</i>	<i>recall</i>
50	0.001	0.58	1	0.06	0.66	0.66	0.28
	0.01	0.38	0.38	1	0.72	0	0
	0.1	0.33	0.33	1	0.38	0.38	1
200	0.001	0.72	0	0	0.63	0.5	0.07
	0.01	0.55	0	0	0.38	0.38	1
	0.1	0.47	0.47	1	0.38	0.38	1
500	0.001	0.77	0.6	0.6	0.88	0.81	0.92
	0.01	0.44	0.44	1	0.33	0.33	1
	0.1	0.27	0.27	1	0.61	0	0
1000	0.001	0.77	0.64	0.75	0.91	0.85	0.92
	0.01	0.5	0.43	1	0.36	0.36	1
	0.1	0.3	0.3	1	0.44	0.44	1
5000	0.001	0.83	0.66	0.66	0.88	0.77	0.77
	0.01	0.63	0.53	1	0.44	0.44	1
	0.1	0.36	0.36	1	0.38	0.38	1

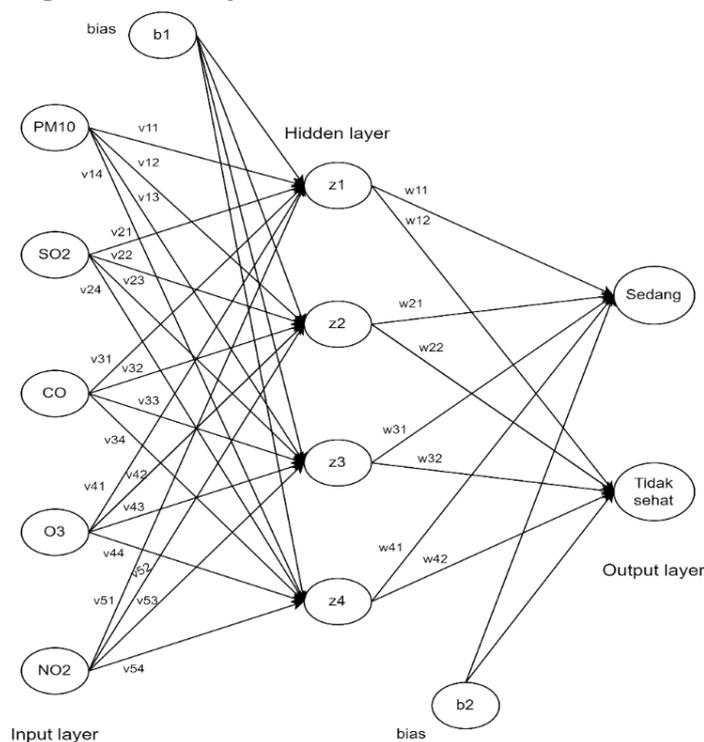
**Tabel 6.** Perhitungan hasil dengan *hidden layer* 3 dan 4

		<i>hidden layer</i>					
		<b>3</b>			<b>4</b>		
<i>epoch</i>	<i>LR</i>	<i>accuracy</i>	<i>precision</i>	<i>recall</i>	<i>accuracy</i>	<i>precision</i>	<i>recall</i>
50	0.001	0.58	0	0	0.66	0	0
	0.01	0.694	0	0	0.47	0.47	1
	0.1	0.61	0	0	0.58	0	0
200	0.001	0.63	1	0.071	0.69	0.66	0.53
	0.01	0.25	0.25	1	0.66	0	0
	0.1	0.33	0.33	1	0.36	0.36	1
500	0.001	0.805	0.75	0.54	0.75	0.53	0.7
	0.01	0.44	0.44	1	0.69	0	0
	0.1	0.36	0.36	1	0.47	0.47	1
1000	0.001	0.83	0.9	0.66	0.86	0.9	0.71
	0.01	0.33	0.33	1	0.38	0.38	1
	0.1	0.61	0	0	0.63	0	0
5000	0.001	0.861	0.69	0.9	0.94	0.9	1
	0.01	0.638	0.5	1	0.61	0.53	1
	0.1	0.36	0.36	1	0.36	0.36	1

**Tabel 7.** Perhitungan hasil dengan *hidden layer* 5

		<i>hidden layer</i>		
		<b>5</b>		
<i>epoch</i>	<i>LR</i>	<i>accuracy</i>	<i>precision</i>	<i>recall</i>
50	0.001	0.77	0.5	0.25
	0.01	0.25	0.25	1
	0.1	0.72	0	0
200	0.001	0.83	0.72	0.72
	0.01	0.361	0.361	1
	0.1	0.5	0	0
500	0.001	0.75	0.58	0.63
	0.01	0.611	0	0
	0.1	0.41	0.41	1
1000	0.001	0.69	0.63	0.5
	0.01	0.41	0.41	1
	0.1	0.38	0.38	1
5000	0.001	0.88	1	0.69
	0.01	0.861	0.8181	0.75
	0.1	0.69	0	0

Setelah dilakukan perhitungan, maka dapat diketahui arsitektur jaringan yang paling optimal dalam mengklasifikasi kualitas udara adalah arsitektur jaringan yang terdiri dari 5 *input layer*, 4 *hidden layer*, dan 2 *output layer* dengan nilai *epoch* 5000 dan *learning rate* 0.001. Arsitektur jaringan dapat dilihat sebagai berikut:



**Gambar 1.** Arsitektur Jaringan

### **Confusion Matrix**

Confusion matrix akan disajikan berdasarkan hasil klasifikasi pada proses *testing*. *Output* hasil klasifikasi terdiri dari nilai 0 yang berarti termasuk kedalam kategori udara sedang dan nilai 1 yang berarti termasuk kedalam kategori udara tidak sehat. Berikut adalah hasil *confusion matrix* menggunakan data *testing* dengan arsitektur jaringan yang terdiri dari 5 *input layer*, 4 *hidden layer*, dan 2 *output layer*:

**Tabel 8.** Confusion Matrix

		Nilai Prediksi	
		Sedang	Tidak sehat
Nilai Aktual	Sedang	20	0
	Tidak sehat	2	14

Berdasarkan hasil *confusion matrix* diatas dapat dihitung nilai *accuracy*, *precision*, dan *recall* menggunakan persamaan:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% = \frac{20+14}{20+14+2+0} \times 100\% = 94\%$$

$$\text{Precision} = \frac{TP}{TP+FP} \times 100\% = \frac{20}{20+2} \times 100\% = 90\%$$

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\% = \frac{20}{20+0} \times 100\% = 100\%$$

Berdasarkan perhitungan dengan arsitektur jaringan dengan 5 *input layer*, 4 *hidden layer*, 2 *output layer*, nilai *epoch* 5000, dan *learning rate* 0.001 didapatkan nilai *accuracy* sebesar 94%, *precision* 90%, dan *recall* 100%. Nilai *accuracy* sebesar 94% menandakan bahwa

model sangat baik dalam mengklasifikasi data (*excellent classification*). Nilai *precision* sebesar 90% menggambarkan tingkat keakuratan antara data yang diminta dengan nilai prediksi yang diberikan oleh model. Nilai *recall* sebesar 100% menggambarkan tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi.

#### D. Kesimpulan

Berdasarkan hasil analisis dalam mengklasifikasi kualitas udara menggunakan metode *Artificial Neural Network* (ANN) dengan algoritma *backpropagation* didapat arsitektur jaringan yang optimal yang terdiri dari 5 *input layer*, 4 *hidden layer*, dan 2 *output layer* dengan nilai *epoch* sebesar 5000 dan *learning rate* 0.001. Arsitektur tersebut menghasilkan nilai *accuracy* yang diperoleh sebesar 94%, *precision* 90%, dan *recall* 100%. Sehingga dapat disimpulkan bahwa metode *Artificial Neural Network* (ANN) algoritma *backpropagation* dalam mengklasifikasi kualitas udara di Provinsi DKI Jakarta pada Tahun 2021 menggunakan fungsi aktivasi Tanh sudah sangat baik dalam melakukan klasifikasi.

#### Acknowledge

Penulis ingin mengucapkan terima kasih sebanyak-banyaknya kepada:

1. Allah SWT atas rahmat dan hidayah-Nya, penulis dapat menyelesaikan skripsi ini dengan baik.
2. Keluarga yang senantiasa selalu mendoakan dan memberikan dukungan kepada penulis.
3. Bapak Abdul Kudus, S.Si., M.Si., Ph.D. selaku Dekan FMIPA Unisba.
4. Bapak Dr. Aceng Komarudin Mutaqin, S.Si., M.T., M.Si. selaku Ketua Prodi Statistika Unisba
5. Bapak Dr. Suwanda, M.S. selaku Dosen Pembimbing yang telah memberikan pengarahan, bimbingan, dan saran kepada penulis dalam penulisan skripsi ini.
6. Bapak Dr. Nusar Hajarisman, M.S. dan Ibu Marizsa Herlina, S.Stat., M.Sc. selaku Dosen Penguji yang telah memberikan masukan dan saran untuk kesempurnaan skripsi ini.
7. Risti, Ayun, Equin, Rengga, Yassar, dan Rizqi yang telah membantu dan memberikan dukungan kepada penulis selama melakukan penulisan skripsi ini.
8. Aulia, Widhi, Arik, Desi, dan Khalda yang membersamai penulis dalam melakukan penulisan skripsi dan selama perkuliahan.
9. Seluruh Dosen Prodi Statistika Unisba yang telah memberikan ilmu pengetahuan selama perkuliahan.
10. Semua pihak lainnya yang telah terlibat dalam penulisan proposal skripsi ini dan tidak dapat penulis sebutkan satu per satu.

#### Daftar Pustaka

- [1] Ordukaya, E., & Karlik, B. (2017). Quality Control of Olive Oils Using Machine Learning and Electronic Nose. *Journal of Food Quality*, 2017.
- [2] Effendi, A. (2013). Penggunaan Artificial Neural Network untuk Mendeteksi Kelainan Mata Miopi pada Manusia dengan Metode Backpropagation.
- [3] Tambun, M. S., Furqon, M. T., & Widodo, A. W. (2018). Penerapan Algoritme Jaringan Syaraf Tiruan Backpropagation pada Pengklasifikasian Status Gizi Balita.
- [4] Aini, D. H. N. (2022). Implementasi Artificial Neural Network dengan Algoritma Backpropagation untuk Klasifikasi Penilaian Rating pada Penjualan Blackmores di Tokopedia. *Braz Dent J.*, 33(1), 1–12.
- [5] Nadhira. (2018). Algoritma Backpropogation Klasifikasi Artikel Ekonomi Fashion dan Ekonomi Film.
- [6] Novita, D. D., Sesunan, A. B., Telaumbanua, M., Triyono, S., & Saputra, T. W. (2021). Identifikasi Jenis Kopi Menggunakan Sensor E-Nose Dengan Metode Pembelajaran Jaringan Syaraf Tiruan Backpropagation. *Jurnal Ilmiah Rekayasa Pertanian Dan Biosistem*.
- [7] Al Rivian, M. E., Rachmat, N., & Rizki Ayustin, M. (2020). Klasifikasi Jenis Kacang-Kacangan Berdasarkan Tekstur Menggunakan Jaringan Syaraf Tiruan. *Jurnal Komputer*

- Terapan.
- [8] Kusnandar, M. (2020). Permen LHK Nomor 14 Tahun 2020. Permen LHK Nomor 14 Tahun 2020 Tentang Indeks Standar Pencemar Udara (ISPU).
  - [9] Yunial, agus heri. (2020). Analisa Perbandingan Algoritma Klasifikasi Support Vector Machine, Decession Tree Dan Naive Bayes. Prosiding Seminar Informatika Dan Sistem Informasi, 5(2), 138–156.
  - [10] Julpan, Nababan, E. B., & Zarlis, M. (2015). Analisis Fungsi Aktivasi Sigmoid Biner dan Sigmoid Bipolar dalam Algoritma Backpropagation pada Prediksi Kemampuan Siswa. *Jurnal Teknovasi*, 02, 103–116.
  - [11] Retnowardhani, Astari & Ramdani, Tiswa. (2019). “Apakah Deep Learning?”. Binus University, [mmsi.binus.ac.id/2019/11/26/apakah-deep-learning/](https://mmsi.binus.ac.id/2019/11/26/apakah-deep-learning/). Diakses pada 29 Mei 2023.
  - [12] Perdana, Rizal Setya. (2017). Pengukuran Akurasi Menggunakan Precision dan Recall. Medium, <https://rizalespe.com/pengukuran-akurasi-menggunakan-precision-dan-recall-71c04988e6ab>. Diakses pada 6 Juli 2023.
  - [13] Utami, Aprilia T.W; Ulama, B. S. S. (2015). Penerapan Backpropagation untuk Meningkatkan Efektivitas Waktu dan Akurasi pada Data. 4(2).
  - [14] Bohari, A. R. (2022). Meningkatkan Kinerja Backpropagation Neural Network Menggunakan Algoritma Adaptif. *IMTechno: Journal of Industrial Management and Technology*.
  - [15] Brian, T. (2017). Analisis Learning Rates Pada Algoritma Backpropagation Untuk Klasifikasi Penyakit Diabetes. *Edutic - Scientific Journal of Informatics Education*, 3(1).
  - [16] Sari, E. Y., Kusri, & Sutoyo, A. (2019). Analisis Jaringan Syaraf Tiruan dengan Backpropagation untuk Prediksi Mahasiswa Dropout. *Citec Journal*, 6(2).
  - [17] Muhammad Rizq Nafisyah Alam, & Aceng Komarudin Mutaqin. (2023). Pemodelan Distribusi Poisson-Sujatha pada Data Frekuensi Klaim Asuransi Kendaraan Bermotor di Indonesia. *Jurnal Riset Statistika*, 71–78. <https://doi.org/10.29313/jrs.v3i1.1944>