

# Air Quality Index Classification Using Neural Network Algorithms

<sup>1</sup>Ela Nurelasari, <sup>2</sup>Esty Purwaningsih

<sup>1,2</sup>Program Studi Sistem Informasi, Universitas Bina Sarana Informatika

Email: ela.eur@bsi.ac.id

## Abstract

*Water has a very important role in maintaining the survival of living things on earth. DKI Jakarta is the capital city of the Republic of Indonesia whose population is categorized as high and is always increasing and is not proportional to the growth of green open spaces and the preservation of green areas, especially in urban areas which are not matched by the size of the area. The level of density of motorized vehicles and infrastructure development is increasing, which can affect air quality. This can result in higher levels of air pollution. The focus of this research is to analyze the air quality index classification in DKI Jakarta using the Neural Network method. The data used is secondary data with variables as determinants consisting of 5 variables, namely PM10, SO2, CO, O3, NO2 and 1 output variable, namely good, moderate, unhealthy and very unhealthy. From the calculation results in this study it is known that the Neural Network method obtains an accuracy of performance of 97.76%*

**Keywords:** Data Mining, Air Pollution, Neural Network Method

## 1. INTRODUCTION

Udara mempunyai peranan yang sangat penting dalam menjaga kelangsungan hidup makhluk hidup yang ada di bumi [1]. DKI Jakarta merupakan Ibu Kota Negara Republik Indonesia yang jumlah penduduknya dikategorikan tinggi dan selalu meningkat dan tidak sebanding dengan pertumbuhan ruang terbuka hijau dan pelestarian kawasan hijau khususnya pada wilayah perkotaan yang tidak diimbangi dengan luas wilayah. Tingkat kepadatan kendaraan bermotor dan pembangunan infrastruktur semakin meningkat, dapat mempengaruhi terhadap kualitas udara. Hal tersebut dapat berakibat tingkat pencemaran udara semakin tinggi. Berdasarkan situs pengamat kualitas udara aqicn.org tanggal 28 Juli 2022 pukul 11.42 , Indonesia menempati urutan ke 25 dari 100 negara teratas dengan polusi udara terburuk [2]. Kualitas udara di DKI Jakarta mencapai level 147 *Unhealthy for Sensitive Groups US Air Index Quality* (AIQ) yang mengindikasikan udara di DKI Jakarta tidak sehat. Dimana *Air Index Quality*(AQI) masuk dalam range 101-150 , tingkat polusi udaranya tidak sehat untuk group sensitive , implikasi kesehatannya untuk anggota kelompok yang sensitive mungkin akan mengalami efek kesehatan, tapi untuk masyarakat umum tidak mungkin terpengaruh. Anggota kelompok sensitive itu termasuk anak-anak dan orang dewasa yang aktif dan orang dengan penyakit pernapasan seperti asma, harus membatasi aktivitas di luar ruangan yang berkepanjangan. Kualitas udara dengan kategori tidak sehat dapat membuat setiap orang mengalami efek kesehatan, bahkan orang yang memiliki riwayat

penyakit pernapasan dapat menyebabkan efek kesehatan yang lebih serius [3]. Timbulnya penyakit kritis dan kronis bagi manusia salah satunya bisa disebabkan oleh kualitas udara yang buruk [4]. Pemerintah Indonesia berdasarkan Keputusan Badan Pengendalian Dampak Lingkungan (Bapedal) Nomor KEP-107/Kabapedal/11/1997 menetapkan pedoman teknis perhitungan dan pelaporan serta informasi indeks standar pencemaran udara (ISPU) [5] untuk menentukan kualitas udara pada suatu daerah dan bagaimana dampaknya terhadap kesehatan serta parameter-parameter dasar untuk indeks standar pencemaran udara dan periode waktu pengukuran.

Baku mutu udara ambien nasional ditetapkan sebagai batas maksimum mutu udara ambien untuk mencegah terjadinya pencemaran udara [6] sebagaimana terlampir dalam PP No 41 Tahun 1999. Pemerintah menetapkan Baku Mutu Udara Ambien Nasional untuk melindungi kesehatan dan kenyamanan masyarakat. Baku Mutu Udara Ambien Nasional dilihat pada Tabel 1.

**Tabel 1 Baku Mutu Udara ambien nasional menurut PP No. 41 tahun 1999**

No	Parameter	Waktu	Baku Mutu
1	Aerosol/Partikulat(PM10)	24 Jam	150 $\mu\text{g}/\text{m}^3$
2	Karbonmonoksida(CO)	1 Jam	30000 $\mu\text{g}/\text{m}^3$
		24 Jam	10000 $\mu\text{g}/\text{m}^3$
3	Ozon(O3)	1 Jam	235 $\mu\text{g}/\text{m}^3$
		1 Tahun	50 $\mu\text{g}/\text{m}^3$
4	Sulfurdioksida(SO2)	24 Jam	365 $\mu\text{g}/\text{m}^3$
		1 Tahun	80 $\mu\text{g}/\text{m}^3$
5	Nitrogendioksida(NO2)	1 Jam	0,25 $\mu\text{g}/\text{m}^3$
		1 Tahun	100 $\mu\text{g}/\text{m}^3$

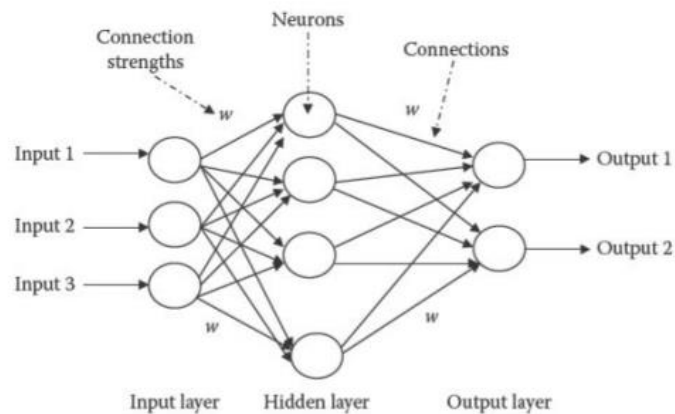
**Tabel 2 kategorikal Atribut yang digunakan**

No	Atribut	Nilai	Kategori
1	Aerosol/Partikulat(PM10)	0	1-150
		1	>150
2	Karbonmonoksida(CO)	0	1-10000
		1	>10000
3	Ozon(O3)	0	1-50
		1	>50
4	Sulfurdioksida(SO2)	0	1-80
		1	>80

5	Nitrogendioksida(NO <sub>2</sub> )	0	1-100
		1	>100

Pada penelitian [7] bahwa penelitian tentang Polusi udara yang terjadi di Cina Utara pada periode januari-oktober 2006 dilakukan dengan menggunakan model MM5, model WRF, dan model ansambel. Ditemukan bahwa model ansambel memiliki akurasi simulasi yang lebih tinggi daripada MM5 dan model WRF. Hasil menunjukkan bahwa kinerja model ansambelCMAQ ditingkatkan oleh 7,0% dan 11,1% serta oleh 17,8% dan 27,5% bila dibandingkan dengan MM5eCMAQ dan model WRFeCMAQ. Umumnya, model ansambel berbasis jaringan saraf adalah alternatif yang efektif untuk mengintegrasikan hasil pemodelan Meteorologi, dan dikaitkan dengan biaya komputasi kurang dibandingkan dengan model Meteorologi numerik. Dapat diharapkan bahwa pendekatan berbasis jaringan neural yang diusulkan juga dapat diterapkan ke daerah lain untuk meningkatkan efektivitas manajemen kualitas udara. Penelitian sebelumnya yang dilakukan oleh [8] yang meneliti tentang penerapan metode klasifikasi support vector machine (SVM) pada data akreditasi Sekolah Dasar (SD) di Kabupaten Magelang menerapkan metode klasifikasi support vector machine (SVM) pada data akreditasi Sekolah Dasar (SD) di Kabupaten Magelang. Hasil yang didapat dari penelitian tersebut yaitu akurasi klasifikasi terbaik yaitu menggunakan fungsi kernel Gaussian Radial Basic Function (RBF) karena menghasilkan akurasi yang lebih besar. Berdasarkan uraian di atas, penelitian ini menggunakan metode Neural Network untuk memprediksi kualitas udara di Provinsi DKI Jakarta yang diimplementasikan pada sistem prediksi kualitas udara (SPKU) DKI Jakarta.

Neural network adalah suatu sistem pemroses informasi yang memiliki karakteristik menyerupai dengan jaringan saraf biologi pada manusia. Neural network didefinisikan sebagai sistem komputasi di mana arsitektur dan operasi diilhami dari pengetahuan tentang sel saraf biologis di dalam otak, yang merupakan salah satu representasi buatan dari otak manusia yang selalu mencoba menstimulasi proses pembelajaran pada otak manusia tersebut [9]. Neural Network (NN) atau Jaringan Syaraf Tiruan (JST) adalah suatu usaha untuk meniru fungsi otak manusia. Otak manusia diyakini terdiri dari jutaan unit pengolahan kecil, yang disebut neuron, yang bekerja secara paralel. Neuron saling terhubung satu sama lain melalui koneksi neuron. setiap individu neuron mengambil input dari satu set neuron. Ini kemudian memproses input tersebut dan melewati output untuk satu set neuron. Keluaran dikumpulkan oleh neuron lain untuk diproses lebih lanjut. Para otak manusia adalah jaringan kompleks neuron dimana koneksi tetap melanggar dan membentuk. Banyak model mirip dengan otak manusia telah diusulkan [10].



**Figure 1. Arsitektur Artificial Neural Network [10]**

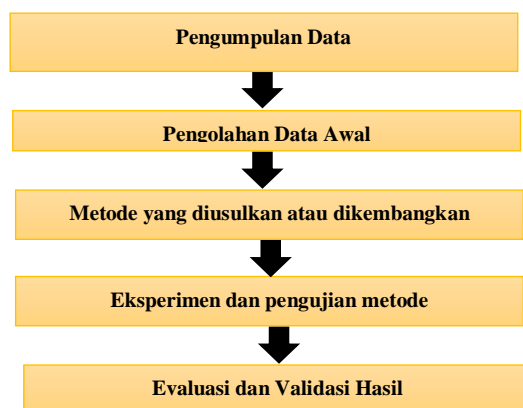
Neural network terdiri dari kumpulan node (neuron) dan relasi. Ada tiga tipe node (neuron) yaitu, input, hidden dan output. Setiap relasi menghubungkan dua buah node dengan bobot tertentu dan juga terdapat arah yang menunjukkan aliran data dalam proses[11]. Algoritma pelatihan backpropagation terdiri dari dua tahapan yaitu feedforward dan backpropagation dari galatnya. Langkah pembelajaran dalam algoritma backpropagation adalah sebagai berikut [12]:

- a. Langkah 1 Inisialisasi bobot jaringan secara acak (biasanya antara -0.1 sampai 0.1)
- b. Langkah 2 Untuk setiap data pada data training, hitung input untuk simpul berdasarkan nilai input dan bobot jaringan saat itu,
- c. Langkah 3 Berdasarkan input yang didapat dari langkah dua, selanjutnya membangkitkan output untuk simpul menggunakan fungsi aktivasi sigmoid: Output
- d. Langkah 4 Hitung nilai error antara nilai yang diprediksi dengan nilai yang sesungguhnya.
- e. Langkah 5 Setelah nilai error dihitung, selanjutnya dibalik ke layer sebelumnya (backpropagation) untuk menghitung nilai error pada hidden layer.
- f. Langkah 6 Nilai Error yang dihasilkan dari langkah sebelumnya digunakan untuk memperbarui bobot relasi.

## 2. METHODS

Pengumpulan data yang digunakan adalah data sekunder yang didapat dari data Badan Pusat Statistik Provinsi DKI Jakarta sebanyak 2000 *record*, dengan variabel sebagai penentu terdiri dari 5 diantaranya pm10, so2, co, o3, no2 dan 1 variabel output yaitu baik, sedang, tidak sehat dan sangat tidak sehat. *Preprocessing* data dilakukan dengan cara menghapus record yang *missing value* serta duplikat. Penelitian adalah sebuah kegiatan yang bertujuan untuk membuat kontribusi orisinal terhadap ilmu pengetahuan [13]. Menurut [13], terdapat empat metode penelitian yang umum digunakan,

diantaranya: action research, experiment, case study dan survey. Penelitian ini menggunakan penelitian eksperimen, yaitu penelitian yang melibatkan penyelidikan perlakuan pada parameter atau variabel tergantung dari penelitiannya dan menggunakan tes yang dikendalikan oleh si peneliti itu sendiri. Dalam penelitian ini juga dilakukan beberapa langkah yang dilakukan dalam metode penelitian seperti pada Gambar berikut:



**Figure 2: Tahapan Penelitian[13]**

Metode Penelitian dibagi menjadi 5 tahapan sebagai berikut[13]:

1. Pengumpulan data (Data Gathering)  
Pada bagian ini dijelaskan tentang bagaimana dan darimana data dalam penelitian ini didapatkan. Pada tahap ini ditentukan data yang akan diproses. Mencari data yang tersedia, memperoleh data tambahan yang dibutuhkan, mengintegrasikan semua data kedalam data set, termasuk variabel yang diperlukan dalam proses.
2. Pengolahan Data Awal (Data Pre-processing)  
Pengolahan data awal meliputi pembersihan data, pentransformasian data ke dalam bentuk yang dibutuhkan serta pengelompokan dan penentuan atribut data.
3. Model atau Metode yang diusulkan atau dikembangkan (Proposed Model/Method) Setelah pengolahan data awal, lalu dibuatkan model yang sesuai dengan jenis data. Pembagian data ke dalam data pelatihan (training data) dan data pengujian (testing data) juga diperlukan untuk pembuatan model.
4. Eksperimen dan Pengujian Model (Method Test and Experiment)  
Metode pengujian model diusulkan pada model yang akan diuji untuk melihat hasil berupa rule yang akan dimanfaatkan dalam mengambil keputusan hasil penelitian. Sebagian data digunakan sebagai data training dan sebagian lagi sebagai data checking. Perhitungan dengan masing-masing algoritma akan diulang beberapa kali untuk mendapatkan besaran parameter terbaik.
5. Evaluasi dan Validasi Hasil (Result Evaluation and Validation)

Pada penelitian ini dilakukan evaluasi terhadap model yang ditetapkan untuk mengetahui tingkat keakurasian model. Evaluasi dilakukan dengan mengamati hasil prediksi menggunakan algoritma soft computing. Validasi dilakukan dengan mengukur hasil prediksi dibandingkan dengan data asal. Pengukuran kinerja dilakukan dengan membandingkan nilai error hasil prediksi.

### 3. RESULTS AND DISCUSSION

Pembuatan model neural network akan dilakukan pada dataset yang terdiri dari 5 atribut yang merupakan atribut dari kualitas polusi udara, dan kelas yang mewakili udara yang baik, sedang, tidak sehat dan sangat tidak sehat. Dari data yang ada hidden layer akan ditentukan 6 hidden layer, model dengan bobot yang sudah diberikan dapat dilihat pada tabel 3.

**Tabel 3. Nilai bobot *hidden layer***

Node	<i>Hidden Layer (Sigmoid)</i>					
	1	2	3	4	5	6
Pm1	0.815	3.245	2.326	14.206	-8.636	9.150
So2	-0.841	-0.866	-5.848	-0.160	1.436	1.775
co	-0.370	-0.811	1.292	0.229	2.124	-2.029
O3	-21.173	18.734	-4.793	-8.707	7.125	-3.978
No2	-0.592	-0.653	-2.591	0.615	1.292	1.441
Bias	-4.039	10.106	-1.087	-5.954	4.564	-4.093

Nilai untuk output dihitung berdasarkan bobot node output dengan nilai dari hidden layer. Setelah diketahui nilai node output, nilai ini dibandingkan dengan nilai output sesungguhnya dan ditambahkan dengan nilai error terdahulu untuk mendapatkan nilai error, setelah didapat nilai error, nilai error ini akan dijumlahkan dengan nilai error sebelumnya dan akan dijadikan untuk menghitung nilai delta. Setelah didapatkan nilai error dan nilai delta, bobot untuk setiap node disesuaikan kembali, Untuk bobot dari attribute R ke hidden layer node pertama, bobot atribut pertama dihitung, nilai perubahan sebelumnya adalah kosong, lalu dihitung nilai perubahan baru .Nilai perubahan ini ditentukan oleh momentum dan learning rate, untuk tahap pertama ini learning rate tidak berpengaruh karena perubahan sebelumnya masih kosong. Nilai perubahan ini ditambahkan dengan nilai bobot sebelumnya untuk menemukan nilai bobot baru. Berikut sampel proses perhitungan untuk hidden1 sebagai berikut:

$$\begin{aligned} \Delta WR_{hidden1} &= (0.1 * 0.001 * 0) + (0.5 * 0) = 0.0000 \Delta WR_{hidden1.new} \\ &= 0.815 + 0.0000 = 0,815 \end{aligned}$$

$$\Delta WR_{hidden2} = (0.1 * 0.001 * 1) + (0.5 * 0) = 0.0001$$

$$\Delta WR_{hidden2.new} = -0.841 + 0.0001 = -0,8409$$

$$\Delta WR \text{ hidden3} = (0.1 * 0.001 * 0) + (0.5 * 0) = 0.0000$$

$$\Delta WR \text{ hidden3.new} = -0.37 + 0.0000 = -0,370$$

$$\Delta WR \text{ hidden4} = (0.1 * 0.001 * 0) + (0.5 * 0) = 0.0000$$

$$\Delta WR \text{ hidden4.new} = -21,173 + 0.0000 = -21,173$$

$$\Delta WR \text{ hidden5} = (0.1 * 0.001 * 0) + (0.5 * 0) = 0.0000$$

$$\Delta WR \text{ hidden5.new} = -0,592 + 0.0000 = -0,592$$

**Tabel 4. Nilai Bobot Akhir untuk Output layer**

Class	Output (Sigmoid)						Threshold
	1	2	3	4	5	6	
Baik	2.853	-7.733	4.824	-7.919	-5.812	-5.531	-1.349
Sedang	13.825	7.943	-5.365	7.260	4.441	-2.136	-14.059
Tidak Sehat	-13.166	-4.914	5.399	1.312	0.117	1.299	5.236
Sangat Tidak Sehat	-2.158	0.134	-3.676	-1.060	0.730	-1.336	-1.416

Hasil *Confusion Matrix*

**Tabel 5. Nilai Akurasi dengan Algoritma Neural Network**

<i>Accuracy: 97,76% +/-0,75% (mikro:97,76%)</i>					
	True Baik	True Sedang	True Tidak Sehat	True Sangat Tidak Sehat	Class precision
<b>Pred. Baik</b>	551	12	0	0	97,87%
<b>Pred. Sedang</b>	12	1807	7	0	98,96%
<b>Pred. Tidak Sehat</b>	0	5	557	31	93,93%
<b>Pred. Sangat Tidak Sehat</b>	0	0	0	7	100%
<b>Class Recall</b>	97,87%	99,07%	98,76%	18,42%	

Metode klasifikasi bisa dievaluasi berdasarkan kriteria seperti tingkat akurasi, kecepatan, kehandalan, skabilitas dan interpretabilitas. Setelah data diolah maka dapat diuji tingkat akurasinya untuk

melihat kinerja dari hasil pengujian diatas, baik evaluasi menggunakan confusion matrix maupun ROC curve terbukti bahwa hasil pengujian algoritma neural network memiliki nilai akurasi yang tinggi sebesar 97,76%.

#### 4. CONCLUSION

Penelitian ini melakukan analisis klasifikasi index kualitas udara di DKI Jakarta dengan menggunakan metode Neural Network. Data yang digunakan adalah data sekunder data Badan Pusat Statistik Provinsi DKI Jakarta sebanyak 2000 *record* dengan variabel sebagai determinan yang terdiri dari 5 variabel yaitu PM10, SO2, CO, O3, NO2 dan 1 variabel keluaran yaitu baik, sedang, tidak sehat dan sangat tidak sehat. Dan Dari hasil perhitungan pada penelitian ini diketahui bahwa metode Neural Network memperoleh performansi akurasi sebesar 97,76%. Dapat disimpulkan bahwa Neural Network dalam klasifikasi Index kualitas udara menunjukkan hasil yang baik dan akurat.

#### REFERENCES

- [1] J. Prayudha, A. Pranata, and A. Al Hafiz, "Implementasi Metode Fuzzy Logic Untuk Sistem Pengukuran Kualitas Udara Di Kota Medan Berbasis Internet of Things (Iot)," *Jurteksi*, vol. 4, no. 2, pp. 141–148, 2018, doi: 10.33330/jurteksi.v4i2.57.
- [2] S. Syihabuddin Azmil Umri, "Analisis Dan Komparasi Algoritma Klasifikasi Dalam Indeks Pencemaran Udara Di Dki Jakarta," *JIKO (Jurnal Inform. dan Komputer)*, vol. 4, no. 2, pp. 98–104, 2021, doi: 10.33387/jiko.v4i2.2871.
- [3] A. Hermawan, "SPKU: Sistem Prediksi Kualitas Udara (Studi Kasus: Dki Jakarta)," 2019, [Online]. Available: <http://eprints.uty.ac.id/3552/>.
- [4] X. Ning, X. Ji, G. Li, and N. Sang, "Ambient PM2.5 causes lung injuries and coupled energy metabolic disorder," *Ecotoxicol. Environ. Saf.*, vol. 170, pp. 620–626, Apr. 2019, doi: 10.1016/J.ECOENV.2018.12.028.
- [5] A. Khumaidi, R. Raafi'udin, and I. P. Solihin, "Pengujian Algoritma Long Short Term Memory untuk Prediksi Kualitas Udara dan Suhu Kota Bandung," *J. Telemat.*, vol. 15, no. 1, pp. 13–18, 2020, [Online]. Available: <https://journal.ithb.ac.id/telematika/article/view/340>.
- [6] A. Kurniawan Stasiun Klimatologi Mlati and B. Meteorologi Klimatologi dan Geofisika, "PENGUKURAN PARAMETER KUALITAS UDARA (CO, NO2, SO2, O3 DAN PM10) DI BUKIT KOTOTABANG BERBASIS ISPU," *J. Teknosains*, vol. 7, no. 1, pp. 1–13, Jul. 2018, doi: 10.22146/TEKNOSAINS.34658.
- [7] S. Cheng, L. Li, D. Chen, and J. Li, "A neural network based ensemble approach for improving the accuracy of meteorological fields used for regional air quality modeling," *J. Environ. Manage.*, vol. 112, pp. 404–414, Dec. 2012, doi: 10.1016/J.JENVMAN.2012.08.020.



- [8] P. A. Octaviani, Y. Wilandari, and D. Ispriyanti, "PENERAPAN METODE KLASIFIKASI SUPPORT VECTOR MACHINE (SVM) PADA DATA AKREDITASI SEKOLAH DASAR (SD) DI KABUPATEN MAGELANG," *J. Gaussian*, vol. 3, no. 4, pp. 811–820, 2014, doi: 10.14710/J.GAUSS.V3I4.8092.
- [9] P. ASTUTI, "KOMPARASI PENERAPAN ALGORITMA C45, KNN DAN NEURAL NETWORK DALAM PROSES KELAYAKAN PENERIMAAN KREDIT KENDARAAN BERMOTOR," *Fakt. Exacta*, vol. 9, no. 1, pp. 87–101, Jun. 2016, Accessed: Sep. 14, 2022..
- [10] R. R. Janghel, A. Shukla, R. Tiwari, and R. Kala, "Breast Cancer diagnosis using Artificial Neural Network models," *Proc. - 3rd Int. Conf. Inf. Sci. Interact. Sci. ICIS 2010*, pp. 89–94, 2010, doi: 10.1109/ICICIS.2010.5534716.
- [11] E. Kusriani, E. T.-Y. A. Offset, and undefined 2009, "Algoritma data mining," *diglib.amikom.ac.id*, Accessed: Sep. 14, 2022.
- [12] "Making Sense of Data: A Practical Guide to Exploratory Data Analysis and ... - Glenn J. Myatt-Google Buku." (accessed Sep. 14, 2022).
- [13] C. W. Dawson, "Projects in Computing and Information Systems A Student's Guide A Student's Guide Second Edition A Student's Guide," Accessed: Sep. 14, 2022. [Online]. Available: [www.pearson-books.com](http://www.pearson-books.com).