

BAB 2

TINJAUAN PUSTAKA

2.1. Pencemaran air

Air merupakan salah satu komponen utama dalam lingkungan yang berperan penting dalam kehidupan (Warlina, 2004). Air, yang bersumber dari sungai, danau, gletser, air hujan, air tanah, dan sumber lainnya, memiliki peran penting, tidak hanya untuk dikonsumsi, tetapi juga digunakan dalam bidang pertanian, peternakan, perhutanan, kegiatan industri, perikanan, dan bidang-bidang lainnya (Effendi, 2015). Karena itu, air harus senantiasa terjaga kualitasnya, untuk mencegah terjadinya pencemaran air, yang dapat membahayakan bagi kehidupan.

Secara harafiah, pencemaran air dapat diartikan sebagai suatu kondisi di mana sebuah kawasan perairan berada dalam kondisi tercemar. Menurut Undang-Undang Nomor 32 Tahun 2009 tentang Perlindungan dan Pengelolaan Lingkungan Hidup, pencemaran lingkungan hidup adalah masuk atau dimasukkannya salah satu atau kombinasi antara makhluk hidup, zat, energi, dan komponen lain ke dalam lingkungan hidup oleh kegiatan manusia sehingga kualitasnya turun sehingga mencapai tingkat tertentu, yang menyebabkan lingkungan hidup tidak dapat berfungsi sesuai dengan peruntukannya. Jadi, dapat disimpulkan bahwa pencemaran air merujuk pada kondisi perairan yang mengandung komponen seperti makhluk hidup atau zat, yang pada kadar tertentu dapat mengganggu keseimbangan ekosistem yang berkembang di dalam kawasan perairan tersebut.

Menurut Warlina (2004), tingkat pencemaran air dapat diukur berdasarkan pengamatan secara fisis, kimiawi dan biologis. Pengamatan secara fisis dapat dilakukan dengan memperhatikan sifat-sifat fisik air, seperti tingkat kejernihan air, perubahan suhu dan warna; serta membandingkan perubahan warna, bau, dan rasa pada air. Peng-

amatan secara kimiawi dapat dilakukan dengan mengukur konsentrasi zat kimia yang terlarut dan tingkat keasaman. Pengamatan secara biologis dapat dilakukan dengan melakukan pengamatan atas mikroorganisme yang terkandung di dalam air, dengan memberikan perhatian khusus pada bakteri yang dapat memicu penyakit.

2.2. Indeks kualitas air (*Water Quality Index*)

Berbagai pendapat telah dikemukakan mengenai indeks kualitas air. Keputusan Menteri Negara Lingkungan Hidup Nomor 115 Tahun 2003 tentang Pedoman Penentuan Status Mutu Air mendefinisikan mutu air sebagai kondisi kualitas air yang diukur dan atau diuji berdasarkan parameter-parameter tertentu dan metode tertentu berdasarkan peraturan perundang-undangan yang berlaku. Efendi (2016) berpendapat bahwa indeks kualitas air adalah sebuah metode untuk mengukur mutu air, yang didapat dari hasil perhitungan parameter-parameter tertentu, di mana parameter tersebut dapat mencerminkan indikator kualitas air.

Penggunaan indeks kualitas air diperkenalkan oleh Horton (1965), di mana kalkulasi indeks kualitas air dilakukan melalui pengukuran parameter seperti *dissolved oxygen* (DO), tingkat keasaman, *coliform*, tingkat kemampuan konduksi, kadar klorin, dan parameter lainnya. Indeks kualitas air juga diperkenalkan oleh Brown *et al.* (1970), dengan menghitung *weight* dari setiap parameter yang diukur.

Menurut Abbasi & Abbasi (2012), indeks kualitas air, di mana setiap parameter dalam pengukurannya ditentukan oleh pihak yang berwenang dalam pengendalian kualitas sumber daya air, dapat memudahkan kegiatan-kegiatan yang membutuhkan pemantauan kualitas air, seperti penelitian sumber daya air, perencanaan lingkungan, pemurnian air, dan kegiatan lainnya. Manfaat dari penerapan indeks kualitas air dalam menentukan kualitas dari sumber daya air adalah sebagai berikut:

- a. Membantu pengambilan keputusan mengenai alokasi dana dan prioritas dalam pengelolaan sumber daya air;
- b. Berperan sebagai media pembanding kualitas sumber daya air antar wilayah geografis tertentu;
- c. Menentukan kesesuaian kualitas sumber daya air di suatu wilayah dengan parameter yang telah ditentukan oleh pihak berwenang;

- d. Memperkirakan perubahan kualitas sumber daya air di suatu wilayah sesuai pergerakan waktu;
- e. Berperan sebagai media informasi kepada publik mengenai tingkatan kualitas sumber daya air di suatu wilayah; dan
- f. Membantu kegiatan penelitian yang memerlukan pengukuran kualitas sumber daya air, seperti dampak aktivitas pembangunan terhadap kualitas sumber daya air, dan penelitian lainnya.

Ukuran kualitas air di Indonesia ditentukan oleh pemerintah melalui Keputusan Menteri Negara Lingkungan Hidup Nomor 115 Tahun 2003 tentang Pedoman Penentuan Status Mutu Air. Penentuan kualitas air dilakukan melalui penghitungan parameter-parameter dengan menggunakan metode Storet. Dalam metode Storet, data dari parameter kualitas air yang terukur akan dibandingkan dengan nilai baku yang disesuaikan dengan tujuan penggunaan air. Sistem nilai yang digunakan dalam menentukan kualitas air adalah sistem nilai dari US-EPA (*Environmental Protection Agency*), di mana kualitas air ditentukan dari empat kelas, yaitu:

- a. kelas A atau baik sekali, dengan skor 0, menunjukkan bahwa sumber daya air yang diukur telah memenuhi nilai baku mutu air;
- b. kelas B atau baik, dengan skor antara -1 hingga -10, menunjukkan bahwa sumber daya air yang diukur berada dalam kondisi tercemar dengan tingkat ringan;
- c. kelas C atau sedang, dengan skor antara -11 hingga -30, menunjukkan bahwa sumber daya air yang diukur berada dalam kondisi tercemar dengan tingkat sedang; dan
- d. kelas D atau buruk, dengan skor kurang dari -31, menunjukkan bahwa sumber daya air yang diukur berada dalam kondisi tercemar dengan tingkat berat.

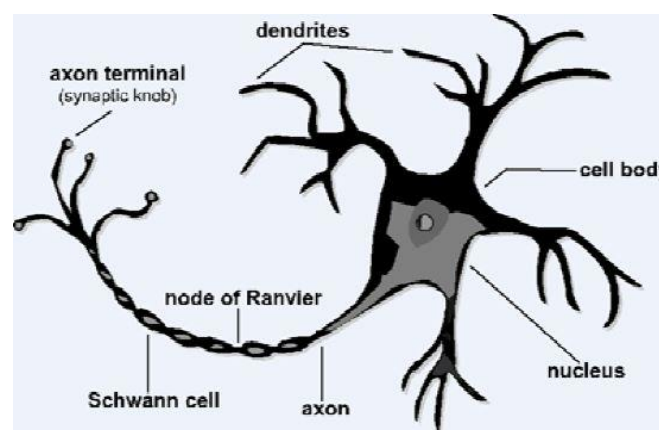
Pengukuran kualitas air menggunakan metode Storet dilakukan dengan mengumpulkan data kualitas air secara berkala. Hasil pengukuran yang didapat akan dibandingkan dengan baku mutu yang dibutuhkan sumber daya air untuk memenuhi tujuan penggunaan. Pengurangan indeks akan dilakukan apabila hasil pengukuran yang didapat tidak sesuai dengan standar baku mutu yang ditentukan. Besar dari pengurangan indeks dijelaskan oleh Tabel 2.1.

Tabel 2.1. Tabel penentuan indeks kualitas air (Sumber: Keputusan Menteri Lingkungan Hidup Nomor 115 Tahun 2003)

Jumlah parameter	Nilai	Parameter		
		Fisika	Kimia	Biologi
Di bawah 10	Maksimum	-1	-2	-3
	Minimum	-1	-2	-3
	Rata-rata	-3	-6	-9
10 ke atas	Maksimum	-2	-4	-6
	Minimum	-2	-4	-6
	Rata-rata	-6	-12	-18

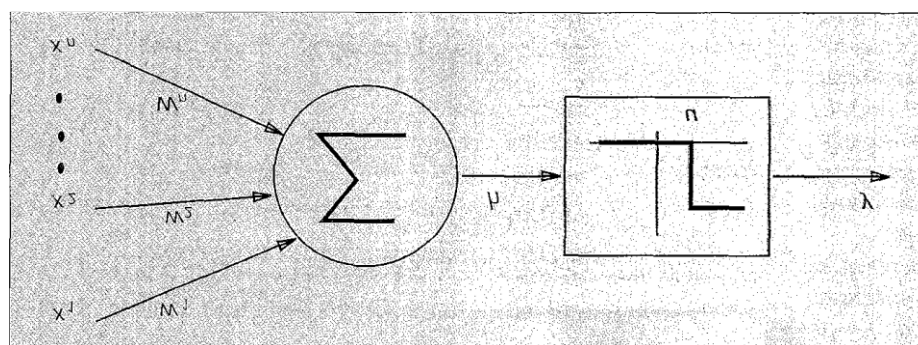
2.3. *Artificial neural networks*

Menurut Hammerstrom (1993), *artificial neural networks* merujuk pada teknologi komputasi yang dikembangkan berdasarkan cara kerja otak pada makhluk hidup. Seperti yang terlihat pada gambar 2.1, otak pada makhluk hidup terdiri atas kumpulan neuron yang saling terhubung, di mana setiap neuron memiliki soma, akson, sinapsis, dan dendrit. Melalui dendrit, sebuah neuron dapat terhubung ke ujung sinapsis dari neuron lainnya. Sinyal yang sampai ke neuron akan diproses di dalam soma sebagai pusat dari neuron, lalu diteruskan lagi melalui dendrit pada ujung lainnya. Dendrit juga dilengkapi oleh akson, yang berfungsi untuk transfer sinyal aktivasi ke neuron lainnya. Menurut Uhrig (1995), sebuah *artificial neural networks* terdiri dari kumpulan unit komputasi yang saling terhubung satu sama lain, melalui koneksi yang memiliki weight tertentu. Unit komputasi ini tersusun dalam rangkaian lapisan (*layer*) tertentu, yang pada umumnya terdiri dari tiga lapisan. Lapisan tersebut adalah *input layer*, *hidden layer*, dan *output layer*.



Gambar 2.1. Struktur neuron pada otak makhluk hidup (Sumber: Heaton, 2008)

Penelitian mengenai *artificial neural networks* dimulai ketika McCulloch & Pitts (1943) mengemukakan struktur neuron untuk melakukan kalkulasi logika. Seperti yang ditunjukkan oleh gambar 2.2, sebuah neuron dapat menerima n sinyal input. Sinyal input yang diterima oleh neuron akan dihitung sesuai fungsi yang berlaku pada neuron, yang kemudian dijumlahkan secara keseluruhan. Hasil penjumlahan yang didapat akan diuji berdasarkan fungsi batasan (*threshold function*). Sebuah neuron dinyatakan dalam kondisi aktif jika hasil penjumlahan sama atau melebihi fungsi batasan tersebut. Sebaliknya, jika hasil penjumlahan tidak melebihi fungsi batasan, maka neuron tersebut dinyatakan dalam kondisi tidak aktif.



Gambar 2.2. Struktur sebuah neuron pada *artificial neural network*

(Sumber: Jain *et al.*, 1996)

Penelitian mengenai peningkatan kinerja pada *artificial neural networks* dimulai ketika Rosenblatt (1958), yang kemudian dikembangkan oleh Rumelhart *et al.* (1986), memperkenalkan *backpropagation neural network*. Dalam *backpropagation neural network*, data yang diterima oleh sebuah node akan dikalkulasikan melalui fungsi aktivasi yang telah ditentukan. Selanjutnya, hasil kalkulasi, beserta nilai *error* yang didapat dari hasil kalkulasi tersebut, akan diteruskan ke neuron lainnya untuk diproses kembali.

2.4. Machine learning

Menurut Hertzmann & Fleet (2011), *machine learning* merujuk pada sebuah metode yang membuat komputer memiliki kemampuan dalam mempelajari dan melakukan sebuah pekerjaan secara otomatis. Proses *machine learning* akan dilakukan melalui

algoritma tertentu, sehingga pekerjaan yang diperintahkan kepada komputer dapat dilakukan secara otomatis.

Machine learning memiliki beberapa pengertian, yaitu sebagai berikut.

- a. *Machine learning* menurut sudut pandang kecerdasan buatan. Proses *learning* merupakan bagian yang penting dalam pembuatan sistem cerdas, namun sistem cerdas tidak dapat dibangun melalui aturan-aturan yang dirancang secara langsung, misalnya dengan menerapkan langkah-langkah prosedural, sehingga dibutuhkan proses yang memungkinkan sistem cerdas untuk mempelajari algoritma secara otomatis.
- b. *Machine learning* menurut sudut pandang teknis. *Machine learning* merupakan sebuah metode yang memungkinkan proses pemrograman komputer melalui sebuah sampel data tertentu.
- c. *Machine learning* menurut sudut pandang statistika. *Machine learning* merupakan suatu metode penyelesaian masalah yang bersifat statistik melalui teknik komputasi tertentu. *Machine learning* menyelesaikan masalah statistik melalui pendekatan yang sedikit berbeda, seperti lebih mengutamakan kecepatan pemrosesan data daripada akurasi hasil pemrosesan.

Machine learning dapat dilakukan melalui 2 fase, yaitu fase *training* dan fase *application*. Dalam fase *training*, model dari algoritma yang digunakan akan dipelajari oleh sistem melalui *training data*. Dalam fase *application*, model yang telah dipelajari sistem melalui fase *training* akan digunakan untuk menghasilkan sebuah keputusan tertentu, dengan menggunakan *testing data*.

Menurut van Heeswijk (2015), *machine learning* dapat dilakukan dengan dua cara, yaitu *supervised learning* dan *unsupervised learning*. Dalam *unsupervised learning*, pemrosesan *sample data* dilakukan tanpa mewajibkan hasil akhir memiliki bentuk yang sesuai dengan bentuk tertentu, dengan menggunakan beberapa *sample data* sekaligus. Penerapan *unsupervised learning* dapat ditemukan pada proses visualisasi, atau eksplorasi data. Sebaliknya, dalam *supervised learning*, *sample data x* akan diproses sedemikian rupa, sehingga menghasilkan bentuk keluaran yang sesuai dengan hasil akhir *y*. *Supervised learning* dapat diterapkan pada proses klasifikasi.

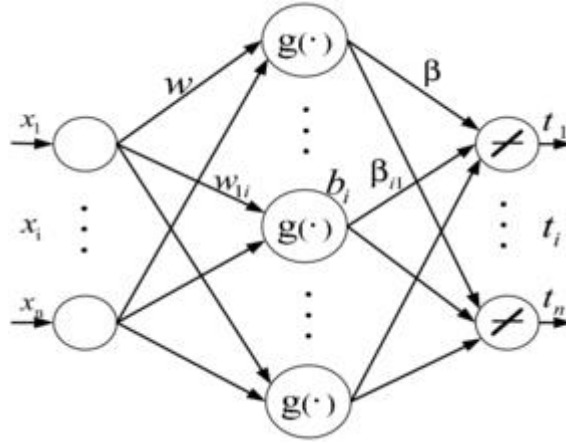
2.5. *Extreme learning machine (ELM)*

Extreme learning machine merupakan sebuah metode yang digunakan dalam jaringan saraf tiruan. Menurut Sun *et al.* (2008), *extreme learning machine* merupakan jaringan saraf tiruan yang bersifat *feedforward* dengan satu *hidden layer*. *Extreme learning machine* dikembangkan oleh Huang *et al.* (2006) untuk memperbaiki permasalahan *learning speed* pada metode-metode yang sebelumnya digunakan pada *feedforward neural networks*. *Feedforward neural networks* memiliki kelemahan dalam *learning speed* karena 2 hal berikut:

- a. Dalam proses *training*, *feedforward neural networks* menggunakan algoritma *slow gradient based learning*; dan
- b. *Input weight* dan *hidden bias* yang berhubungan antar *layer* pada jaringan saraf tiruan ditentukan secara berulang-ulang oleh algoritma *slow gradient based learning*.

Karena kedua hal tersebut, *feedforward neural networks* tidak hanya memiliki *learning speed* yang rendah, namun juga memiliki peluang yang tinggi untuk mencapai sebuah kondisi di mana jaringan saraf tiruan terjebak dalam *local minima* (Huang *et al.*, 2006). Hal ini diperbaiki pada ELM, di mana *input weight* dan *hidden bias* dipilih secara acak, sehingga dapat memperkecil *learning speed* dan bekerja dengan baik.

Struktur umum ELM dapat dilihat pada gambar 2.3. *Extreme learning machine* dijalankan pada *single hidden layer feedforward neural network*, yang digunakan oleh metode ELM. Dalam menjalankan proses *training*, ELM menggunakan teori invers matriks. Teori invers matriks yang digunakan ELM adalah teori *Moore-Penrose pseudoinverse*.



Gambar 2.3. Struktur umum ELM (Sumber: iopscience.iop.org)

Jika diketahui sebuah sampel N berupa kumpulan data (x_i, t_i) , di mana $x_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in R^n$, dan $x_i = [t_{i1}, t_{i2}, \dots, t_{im}]^T \in R^m$, maka sebuah model *single hidden layer feedforward neural network* yang memiliki fungsi aktivasi $g(x)$ dan \bar{N} buah neuron pada *hidden layer*, dapat didefinisikan melalui persamaan 2.1:

$$\sum_{i=1}^{\bar{N}} \beta_i g(w_i \cdot x_j + b_i) = o_j, j = 1, \dots, N \quad (2.1)$$

di mana $w_i = [w_{i1}, w_{i2}, \dots, w_{in}]^T$ merupakan input weight pada vektor yang terhubung antara *hidden neuron* ke- i dan neuron masukan, $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T$ merupakan weight dari vektor yang menghubungkan *hidden neuron* ke- i dan neuron keluaran, dan b_i merupakan nilai *bias* dari *hidden neuron* ke- i . Jika $\sum_{j=1}^N \|o_j - t_j\| = 0$, maka persamaan 2.1 dapat juga didefinisikan melalui persamaan 2.2:

$$\sum_{i=1}^{\bar{N}} \beta_i g(w_i \cdot x_j + b_i) = t_j, j = 1, \dots, N \quad (2.2)$$

Berdasarkan persamaan kedua, maka persamaan N dapat didefinisikan melalui persamaan 2.3:

$$H\beta = T \quad (2.3)$$

di mana tiap elemen H , β , dan T didefinisikan oleh persamaan 2.4, 2.5, dan 2.6:

$$H(w_1, \dots, w_{\bar{N}}, b_1, \dots, b_{\bar{N}}, x_1, \dots, x_N) = \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & \cdots & g(w_{\bar{N}} \cdot x_1 + b_{\bar{N}}) \\ \vdots & \ddots & \vdots \\ g(w_1 \cdot x_N + b_1) & \cdots & g(w_{\bar{N}} \cdot x_N + b_{\bar{N}}) \end{bmatrix} \quad (2.4)$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{\bar{N}}^T \end{bmatrix} \quad (2.5)$$

$$T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix} \quad (2.6)$$

Proses pelatihan (*training*) menggunakan extreme learning machine dimulai dengan memberikan training set $\aleph = \{(x_i, t_i) | x_i \in R^n, t_i \in R^m, i = 1, \dots, N\}$, fungsi aktivasi $g(x)$, dan \bar{N} sebagai jumlah *hidden neuron*. Langkah-langkah yang dilakukan dalam proses training menggunakan extreme learning machine adalah sebagai berikut:

- Menentukan *input weight* w_i dan *bias* b_i secara acak;
- Menghitung nilai dari matriks *hidden layer output* H ; dan
- Menghitung nilai dari matriks keluaran β , yang ditentukan oleh persamaan 2.7:

$$\beta = H^\dagger T \quad (2.7)$$

di mana H^\dagger merupakan hasil *pseudoinvers* dari matriks H , dengan H^\dagger adalah matriks G yang memenuhi persamaan 2.8:

$$AGA = A, GAG = G, (AG)^T = AG, (GA)^T = GA \quad (2.8)$$

2.6. Penelitian Terdahulu

Penelitian mengenai analisis kualitas lingkungan, terutama kualitas udara dan kualitas air, telah dilakukan oleh beberapa orang. Ayyalasomayajula *et al.* (2016) melakukan simulasi kualitas udara menggunakan *big data* melalui Apache Hadoop. Hasil penelitian menunjukkan peningkatan kinerja sebesar 20 hingga 25 % dalam proses simulasi data. Khan & Chai (2016) melakukan prediksi tingkat pencemaran air menggunakan *artificial neural network*. *Artificial neural network* yang digunakan berupa *perceptron* dengan satu *hidden layer*. Qiao *et al.* (2008) melakukan evaluasi kualitas air di Sungai Changjiang menggunakan *fuzzy logic*. Ming (2012) melakukan asesmen kualitas air menggunakan *backpropagation neural network*. Pengujian dilakukan terhadap 10 sampel, di mana nilai *error* semakin mendekati 0 untuk setiap pengulangan.

Selain itu, penelitian mengenai penerapan *extreme learning machine* (ELM) sudah dilakukan dalam berbagai bidang. Huang *et al.* (2013) menggunakan *extreme learning machine* untuk mendeteksi tumor liver berdasarkan citra yang ditangkap dari proses *CT scan*. Penelitian ini memberikan hasil deteksi dengan volumetric overlap error sebesar 67,15 %. Pangaribuan & Suharjito (2014) menggunakan *extreme learning machine* untuk diagnosis diabetes mellitus. Dalam penelitian ini, proses *training* membutuhkan waktu rata-rata 0,1687 detik, dibandingkan dengan 0,9040 detik dengan menggunakan *backpropagation*. Selain itu, *error rate* yang dihasilkan mencapai 0,4036, dibandingkan 0,9425 dengan menggunakan *backpropagation*.

Pada penelitian ini, proses prediksi dilakukan menggunakan *extreme learning machine*. Selain itu, data yang digunakan tidak disimpan melalui kluster tertentu, melainkan menggunakan *file* data tersendiri. Data tersebut akan diproses menggunakan *extreme learning machine*, sehingga menghasilkan grafik yang menggambarkan perkiraan tingkat kualitas air. Rincian singkat dari penelitian yang pernah dilakukan dapat dilihat pada Tabel 2.2.

Tabel 2.2. Penelitian Terdahulu

No.	Peneliti	Tahun	Judul	Keterangan
1	Ayyalasomayajula <i>et al.</i>	2016	Air quality simulations using big data programming models	Peningkatan kinerja sebesar 20-25 %
2	Khan & Chai	2016	Predicting and analyzing water quality using machine learning: a comprehensive model	Training error: 0,0192 hingga 0,0444, testing error: 0,0181 hingga 0,0470
3	Qiao <i>et al.</i>	2008	The application of fuzzy comprehensive evaluation on the water quality of Changjiang river	-
4	Ming	2012	A New Water Quality Assessment Method Based on BP Neural Network	<i>Error rate</i> semakin mendekati 0 untuk setiap pengulangan
5	Huang <i>et al.</i>	2013	Liver tumor detection using kernel-based <i>extreme learning machine</i>	Volumetric overlap error rata-rata 67,15 %
6	Pangaribuan & Suharjito	2014	Diagnosis of diabetes mellitus using <i>extreme learning machine</i>	<i>Error rate</i> 0,4036 dan kecepatan training rata-rata 0,1687 detik