

BAB 3

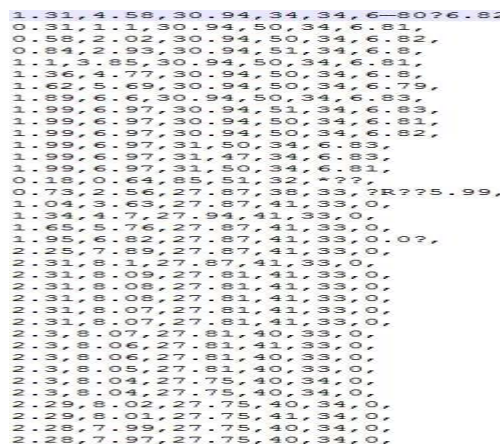
ANALISIS DAN PERANCANGAN SISTEM

Bab ini membahas tentang implementasi *extreme learning machine* dalam proses klasifikasi kualitas air di Danau Toba. Bab ini juga membahas tentang data yang digunakan serta proses normalisasi data. Selain itu, bab ini juga membahas tentang proses training, testing, dan validasi dari keluaran yang dihasilkan berdasarkan data yang dihasilkan dari proses normalisasi data.

3.1. Data yang Digunakan

Data yang digunakan dalam penelitian ini memiliki format CSV (*Comma Separated Value*), di mana nilai dari setiap parameter dipisahkan oleh sebuah karakter *semicolon* (titik koma), seperti yang ditunjukkan pada gambar 3.1. Data ini didapat dari penelitian yang dilakukan oleh Rahmat *et al.* (2016) mengenai pengukuran kualitas air di Danau Toba. Pengukuran dilakukan di empat lokasi, yaitu:

- Haranggaol Horison, Kabupaten Simalungun;
- Ambarita, Kabupaten Samosir;
- Ajibata, Kabupaten Toba Samosir; dan
- Parapat, Kabupaten Simalungun.



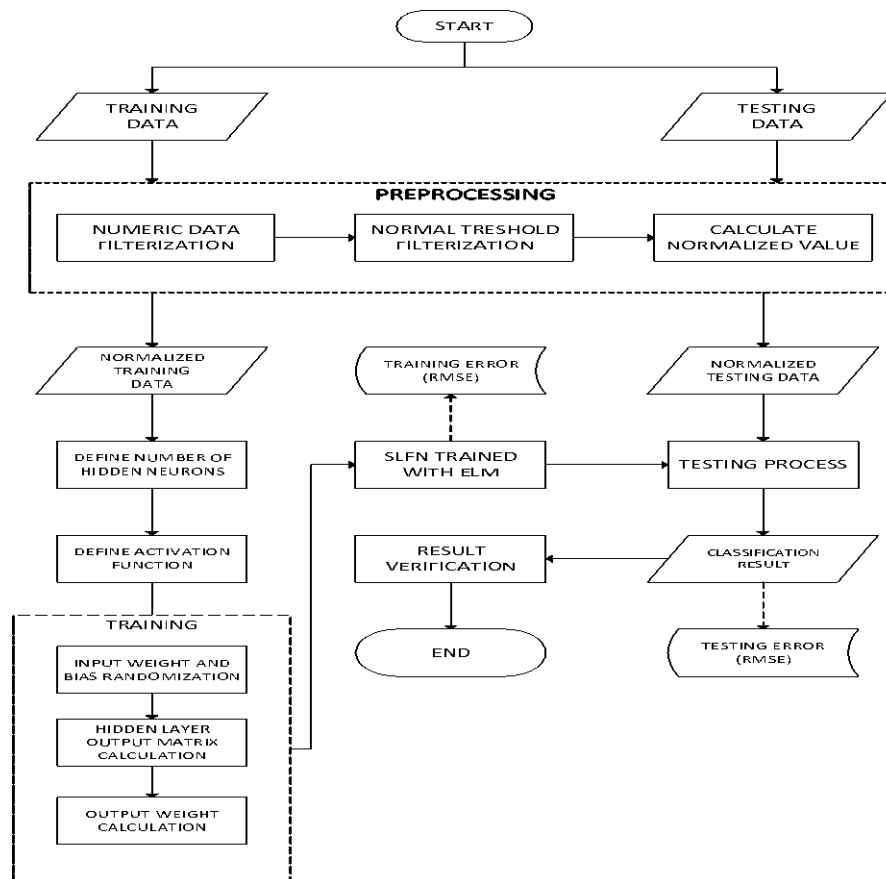
```
1.31,4.58,30.94,34.34,68076.82,
0.31,1.1,30.94,50.34,6.81,
0.58,2.02,30.94,50.34,6.82,
0.84,2.93,30.94,51.34,6.8,
1.1,3.85,30.94,50.34,6.81,
1.36,4.77,30.94,50.34,6.8,
1.62,5.69,30.94,50.34,6.79,
1.89,6.6,30.94,50.34,6.83,
1.99,6.97,30.94,51.34,6.83,
1.99,6.97,30.94,50.34,6.81,
1.99,6.97,30.94,50.34,6.82,
1.99,6.97,31.50,34.6.83,
1.99,6.97,31.47,34.6.83,
1.99,6.97,31.50,34.6.81,
0.18,0.64,85.51,32.77,
0.73,2.56,27.87,38.33,7R775.99,
1.04,3.63,27.87,41.33,0,
1.34,4.7,27.94,41.33,0,
1.65,5.76,27.87,41.33,0,
1.95,6.82,27.87,41.33,0.02,
NNN.25,7.89,27.87,41.33,0,0,
NNN.31,8.1,27.87,41.33,0,
NNN.31,8.09,27.81,41.33,0,
NNN.31,8.08,27.81,41.33,0,
NNN.31,8.08,27.81,41.33,0,
NNN.31,8.07,27.81,41.33,0,
NNN.31,8.07,27.81,41.33,0,
NNN.3,8.07,27.81,40.33,0,
NNN.3,8.06,27.81,41.33,0,
NNN.3,8.06,27.81,40.33,0,
NNN.3,8.05,27.81,40.33,0,
NNN.3,8.04,27.75,40.34,0,
NNN.3,8.04,27.75,40.34,0,
NNN.3,8.02,27.75,40.34,0,
NNN.3,8.01,27.75,41.34,0,
NNN.28,7.99,27.75,40.34,0,
NNN.28,7.97,27.75,40.34,0,
```

Gambar 3.1. Bentuk dataset yang dihasilkan dari penelitian Rahmat *et al.* (2016)

Data yang digunakan dalam penelitian ini merekam hasil pengukuran kualitas air melalui parameter fisika dan kimia pada air. Parameter tersebut antara lain suhu air, tingkat keasaman air, tingkat oksigen terlarut (*dissolved oxygen*), tingkat potensi reduksi oksidasi (*oxidation reduction potential*), suhu udara, dan kelembaban udara.

3.2. Analisis Sistem

Proses klasifikasi kualitas air Danau Toba berdasarkan data yang diperoleh dari Rahmat *et al.* (2016), dilakukan melalui beberapa langkah. Langkah-langkah tersebut ialah normalisasi data *training* dan data *testing*, penentuan jumlah neuron pada *hidden layer*, penentuan fungsi aktivasi, proses *training*, proses *testing*, dan proses validasi. Setiap langkah yang dilakukan akan dijelaskan dengan lebih terperinci pada bagian-bagian selanjutnya. Adapun arsitektur umum yang menggambarkan metode pada penelitian ini ditunjukkan pada Gambar 3.2.



Gambar 3.2. Arsitektur umum

3.2.1. Preprocessing

Tahap awal dari proses klasifikasi kualitas air Danau Toba pada penelitian ini dimulai dari proses preprocessing. Tahap preprocessing, yang juga dapat disebut sebagai tahap normalisasi, adalah sebuah proses di mana data disesuaikan sehingga memenuhi batasan nilai tertentu. (Shalabi *et al.*, 2006) Hasil dari proses normalisasi adalah kumpulan data yang dapat digunakan pada tahap selanjutnya. Pada penelitian ini, data *training* dan data *testing* yang diterima akan diolah terlebih dahulu, sehingga dihasilkan set data yang dapat digunakan oleh *extreme learning machine*.

Proses preprocessing dilakukan melalui tiga tahap. Tahap pertama dari proses ini adalah melakukan penyaringan terhadap tiap baris data, sehingga hanya data dengan nilai numerik yang akan digunakan untuk proses selanjutnya. Tahap ini dilanjutkan dengan menyaring data sesuai dengan batasan normal setiap parameter, sehingga dihasilkan data yang memiliki nilai yang sesuai dengan batasan setiap parameter. Tahap terakhir dari proses preprocessing adalah menghitung nilai data dengan batasan $[C, D]$. Hasil dari proses preprocessing yang dilakukan adalah sebuah kumpulan data (dataset) yang sesuai dengan batasan yang ditentukan.

Metode normalisasi yang digunakan adalah *min-max normalization*, di mana normalisasi menghasilkan data yang memiliki batasan antara dua nilai tertentu (Patro & Sahu, 2015). Normalisasi akan dilakukan berdasarkan persamaan 3.1:

$$A' = \left(\frac{A - A_{min}}{A_{max} - A_{min}} \right) * (D - C) + C \quad (3.1)$$

di mana A' merupakan hasil normalisasi dari data bernilai A yang memiliki batasan antara $[C, D]$.

3.2.2. Penentuan jumlah neuron pada hidden layer

Menurut Heaton (2008), penentuan jumlah node pada *artificial neural network*, khususnya pada *hidden layer*, merupakan hal yang penting sebelum menjalankan *training*, karena *hidden layer* berperan penting dalam penghitungan hasil akhir dari *artificial neural network*.

Jumlah node pada *hidden layer* yang tidak optimal dapat menyebabkan permasalahan tertentu pada proses *training*. *Hidden layer* yang memiliki jumlah node terlalu sedikit dapat menyebabkan kondisi *underfitting*, di mana node yang tersedia tidak dapat bekerja secara maksimal untuk mendeteksi sinyal yang diterima dari *input*

layer. Sebaliknya, jumlah node yang terlalu banyak dapat berakibat pada bertambahnya waktu yang dibutuhkan *artificial neural network* untuk memproses data. Selain itu, jumlah node yang terlalu banyak juga dapat berakibat pada kondisi overfitting, di mana jumlah informasi yang diterima tidak cukup untuk diproses dalam training karena banyaknya kapasitas pemrosesan informasi yang dimiliki jaringan. Penentuan jumlah neuron pada *hidden layer* dapat memenuhi aturan-aturan berikut:

- a. Jumlah neuron pada *hidden layer* harus melebihi jumlah neuron di *input layer* dan tidak boleh melebihi jumlah neuron di *output layer*;
- b. Jumlah neuron pada *hidden layer* harus setara dengan $2/3$ (dua per tiga) dari jumlah neuron pada *input layer* dan *output layer*; dan
- c. Jumlah neuron pada *hidden layer* tidak boleh melebihi dua kali jumlah neuron pada *input layer*.

Aturan mengenai jumlah neuron pada *hidden layer* ini dapat digunakan sebagai pertimbangan. Walaupun begitu, proses menentukan jumlah neuron pada *hidden layer* merupakan proses *trial and error*. Hal ini dilakukan agar jaringan dapat menyesuaikan diri dengan masalah yang akan diselesaikan.

3.2.3. Penentuan fungsi aktivasi

Tahap yang dilakukan setelah jumlah neuron pada *hidden layer* adalah menentukan fungsi aktivasi yang akan digunakan neuron dalam proses *training* dan proses *testing*.

3.2.4. Proses training