Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 76−84
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5871

Klasifikasi Kelayakan Air Minum dengan Backpropagation Neural Network Berbasis Penanganan Missing Value dan Normalisasi

Saifur Yusuf Kurniawan, Suwanto Sanjaya*, Yelfi Vitriani, Iis Afrianty

Sains dan Teknologi, Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru
Panam, Jl. HR. Soebrantas No.Km. 15, RW.15, Simpang Baru, Kota Pekanbaru, Riau, Indonesia
Email: [1]yusufsicahwa@gmail.com, [2,*]suwantosanjaya@uin-suska.ac.id, [3]yelfi.vitriani@uin-suska.ac.id, [4]iis.afrianty@uin-suska.ac.id
Email Penulis Korespondensi: suwantosanjaya@uin-suska.ac.id
Submitted: **30/08/2024;** Accepted: **05/10/2024;** Published: **14/10/2024**

Abstrak-Permasalahan kualitas air minum yang layak untuk dikonsumsi menjadi isu penting dalam kehidupan sehari-hari, khususnya dalam upaya menjaga kesehatan masyarakat. Penelitian serupa mengenai klasifikasi kelayakan air minum masih belum memberikan hasil yang memuaskan. Tujuan penelitian ini adalah mengklasifikasi data kelayakan air minum menggunakan metode backpropagation neural network guna memastikan air yang dikonsumsi memenuhi standar keamanan. Data yang digunakan adalah data publik dari open repository yang berjumlah 3276 data. Atribut parameter kualitas air berjumlah 9, yaitu pH, Hardness, Solids, Chloramines, Sulfate, Conductivity, Organic carbon, Trihalomethanes, dan Turbidity. Pra-pemrosesan data yang dilakukan adalah penghapusan missing value, mengganti missing value dengan nilai rata-rata atribut, normalisasi menggunakan metode MinMax Scaler dan Z-score. Model arsitektur jaringan syaraf tiruan terdiri dari neuron input, hidden layer, dan output. Hasil skenario arsitektur yang terbaik adalah [9;[17;15;10];1] atau 9 neuron input, 17 neuron pada hidden layer pertama, 15 neuron pada hidden layer kedua, 10 neuron pada hidden layer ketiga dan 1 neuron output. Hasil evaluasi menunjukkan bahwa model ini berhasil mengklasifikasi data kelayakan air minum dengan tingkat akurasi sebesar 0,6579. Hasil penelitian ini menunjukkan akurasi yang dicapai masih perlu ditingkatkan untuk aplikasi yang lebih andal, hasil ini menunjukkan potensi yang baik dari metode BPNN dalam melakukan klasifikasi data kualitas air minum.

Kata Kunci: Akurasi; Backpropagation; Jaringan Syaraf Tiruan; Klasifikasi; Kualitas Air.

Abstract-The issue of drinking water quality and its suitability for human consumption represents a significant concern in contemporary society, particularly in the context of maintaining public health. The existing research on the classification of drinking water eligibility has yet to yield conclusive results. The objective of this research is to utilize the backpropagation neural network method to categorize drinking water feasibility data, thereby ensuring that the water consumed meets established safety standards. The data utilized in this study were obtained from an open repository and encompass a total of 3,276 data points. The data set comprises nine water quality parameter attributes, namely pH, hardness, solids, chloramines, sulfate, conductivity, organic carbon, trihalomethanes, and turbidity. The data underwent a series of pre-processing steps, including the removal of missing values, the replacement of missing values with the average value of the attribute, and normalization using the MinMax Scaler and Z-score methods. The artificial neural network architecture comprises three principal components: input, hidden, and output neurons. The optimal architecture scenario is [9; 17; 15; 10; 1], comprising nine input neurons, 17 neurons in the initial hidden layer, 15 neurons in the second hidden layer, 10 neurons in the third hidden layer, and a single output neuron. The evaluation results demonstrate that this model effectively classifies drinking water eligibility data with an accuracy rate of 0.6579. However, the results indicate that the accuracy achieved requires further improvement for more reliable applications. These findings illustrate the promising potential of the BPNN method in classifying drinking water quality data.

Keywords: Accuracy; Backpropagation; Artificial Neural Network; Classification; Water Quality.

1. PENDAHULUAN

Pada tahun 2021, tercatat 2 miliar lebih orang tinggal di negara yang kekurangan air[1]. Situasi ini diprediksi memburuk disebabkan oleh pertumbuhan populasi dan perubahan iklim[1]. Pada tahun 2022, lebih dari 1,7 miliar orang mengandalkan air minum yang tercemar oleh feses[1]. Masalah kelayakan air minum ini menjadi isu kritis yang memerlukan perhatian serius dari pemerintah, organisasi internasional, dan masyarakat sipil[1]. Dikarenakan Indonesia memiliki potensi sebagai sumber daya air yang sangat signifikan dan berada di peringkat ke-5 dunia, penting untuk menjaga kualitas air agar tetap terjaga. Pada tahun 2022 sebagian besar sungai di negara tersebut masih tercemar berbagai polutan, yang menambah tantangan terhadap kelayakan air minum dunia yang sudah kritis[2]. Sesuai Peraturan Menteri Kesehatan Nomor 492/MENKES/PER/IV/2010, air yang layak diminum harus memenuhi standar kesehatan dan aman diminum secara langsung, sehingga pengelompokkan kualitas sumber air menjadi penting untuk memastikan kualitas air yang layak konsumsi[3].

Penelitian mengenai klasifikasi kelayakan air minum sudah pernah dilakukan sebelumnya. Pada penelitian yang dilakukan Safira, dengan metode jaringan syaraf tiruan (neural network) untuk klasifikasi kualitas air berhasil mencapai akurasi yang sangat baik yaitu 97,6% dengan dataset yang terdiri dari 7999 data dengan 15 atribut dan pembagian 70% untuk data pelatihan, 15% pengujian, dan 15% validasi[4]. Berdasarkan penelitian tersebut, dalam penelitian ini juga akan melanjutkan studi mengenai klasifikasi kelayakan air minum dengan mengimplementasikan metode backpropagation neural network dengan membedakan dataset yang digunakan. Metode Backpropagation Neural Network (BPNN) adalah salah satu metode neural network yang biasa digunakan untuk berbagai keperluan klasifikasi. Metode ini mampu mengklasifikasikan data secara kompleks[5], [6]. Hal ini berkaitan dengan penggunaan arsitektur jaringan syaraf tiruan dengan beberapa lapisan, yakni lapisan input, hidden layer, dan lapisan output[7].

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 76−84
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5871

Kemudian dikombinasikan dengan algoritma backpropagation yakni algoritma pembelajaran terawasi yang menggunakan langkah arah mundur (backward) dalam mengubah nilai bobot- bobotnya[8],[9]. Oleh sebab itu, metode Backpropagation Neural Network menjadi pilihan dalam banyak klasifikasi data dan menjadi solusi untuk klasifikasi data yang efektif[10], [11].

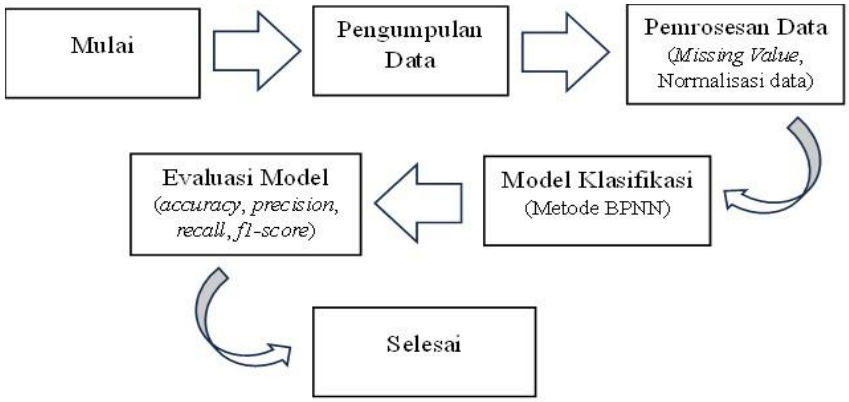
Dataset yang akan digunakan dalam penelitian ini, telah digunakan dalam penelitian terdahulu namun belum menghasilkan akurasi yang memuaskan. Misalnya dalam penelitian Savitri et.al. mendapat kesimpulan bahwa algoritma Random Forest Classifier memiliki akurasi sebesar 76.52% dan itu merupakan hasil yang terbaik dibandingkan dengan beberapa algoritma lainnya yang digunakan, antara lain Logistic Regression, SVM, KNN, dan XGBoost Classifier[12]. Dalam penelitian lain dengan data yang sama namun menggunakan dan membandingkan tiga metode klasifikasi yang berbeda, yaitu K-nearest neighbors, Naïve Bayes, dan Decision Tree, menunjukkan bahwa Decision Tree mencapai tingkat keakuratan tertinggi, yakni sebesar 86,88%[13]. Data tersebut layak digunakan sebagai dataset dalam penelitian ini karena memiliki keragaman atribut yang relevan sejumlah 9 atribut dan jumlah data yang memadai, yaitu 3276 sampel, yang memungkinkan untuk analisis komprehensif dalam klasifikasi kelayakan air minum.

Dalam penelitian yang menggunakan dataset berbeda, sebanyak 226 record dengan atribut turbidity, pH, temperature, sisa khlor, dan total khlor, diperoleh akurasi sebesar 97,35% dengan menggunakan metode Naive Bayes[14]. Studi lain yang mengklasifikasikan kualitas air sumur di Jakarta menggunakan 267 data dengan rasio pembagian data pelatihan dan data pengujian masing-masing 80% dan 20%, menghasilkan presisi sebesar 0,823 dan sensitivitas 0,83 dengan algoritma Random Forest[15]. Klasifikasi kualitas air juga pernah diteliti menggunakan metode Extreme Learning Machine (ELM) dengan hasil nilai evaluasi model untuk data dengan smote memberikan hasil yang paling baik saat menggunakan K-fold=10 dengan tingkat akurasi mencapai 97%[16]. Oleh karena itu, metode neural network memiliki potensi untuk memberikan akurasi yang lebih tinggi dalam klasifikasi kelayakan air minum.

Dengan demikian, penelitian ini dimaksudkan untuk melanjutkan studi mengenai klasifikasi kelayakan air minum dengan menggunakan metode Backpropagation Neural Network (BPNN). Metode BPNN dipilih karena telah terbukti efektif dalam berbagai aplikasi klasifikasi sebelumnya, sedangkan dataset dipilih karena merujuk pada penelitian sebelumnya masih belum menghasilkan akurasi yang memuaskan. Penerapan BPNN pada dataset ini, diharapkan dapat memberikan tingkat akurasi yang memuaskan. Penelitian ini juga diharapkan menjadi solusi yang lebih baik dalam memastikan air yang dikonsumsi dapat diklasifikasikan seseuai standar kualitas yang ditetapkan dan aman untuk digunakan sehari-hari.

2. METODOLOGI PENELITIAN

Metodologi penelitian merupakan penjelasan mengenai langkah-langkah yang diterapkan pada penelitian untuk melakukan penelitian dengan sistematis dan terarah agar penelitian berjalan sesuai dengan tujuan yang dibuat. Tahapan pada penelitian ini dapat dilihat pada gambar 1 berikut:



Gambar 1. Metodologi penelitian

2.1 Pengumpulan Data

Tahapan pengumpulan data dalam penelitian ini dimulai dengan penentuan kebutuhan informasi yang spesifik untuk klasifikasi kelayakan air minum. Data yang digunakan berasal dari sumber sekunder, yakni sumber data yang tidak memberikan data secara langsung kepada pengumpul data atau data ini bersifat publik. Data sekunder dalam sebuah penelitian dapat diperoleh dari jurnal penelitian, buku referensi, internet dan lain – lain[17].

Dalam tahapan ini dilakukan eksplorasi awal pada beberapa situs untuk mengidentifikasi dataset yang relevan dengan atribut yang diperlukan, seperti parameter fisik dan kimia air. Setelah dataset yang tepat ditemukan, dilakukan pengunduhan data untuk selanjutnya data tersebut dievaluasi untuk memastikan kelengkapan dan keandalannya. Evaluasi ini meliputi pengecekan metadata yang menjelaskan sumber data, metode pengumpulan, dan kualitas data[18].

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 76−84
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5871

2.2 Pemrosesan Data

Pemrosesan data adalah tahap permulaan dengan melakukan memodifikasi data yang sudah ditentukan agar siap untuk tahap pengolahan berikutnya. Teknik ini digunakan untuk menangani masalah seperti missing values, data redundan, outliers, atau format data yang tidak sesuai dengan model, karena keberadaan masalah tersebut dapat mengganggu hasil output[19].

Dalam menangani missing value pada dataset, terdapat dua pendekatan umum yang sering digunakan, yaitu penghapusan data dan imputasi rata-rata. Penghapusan data adalah metode sederhana yang melibatkan penghilangan seluruh baris atau kolom yang memiliki nilai yang hilang. Di sisi lain, imputasi rata-rata merupakan metode di mana nilai yang hilang digantikan dengan rata-rata dengan persamaan (1) dari nilai yang tersedia dalam dataset untuk kolom yang bersangkutan. Kedua pendekatan ini menawarkan solusi praktis untuk menangani missing value, dengan pilihan metode yang bergantung pada jumlah data yang hilang dan tujuan analisis data[20].

$$\text{Mean} = \frac{\sum_{i=1}^{n} x_i}{n} \quad (1)$$

Normalisasi data dilakukan dengan membandingkan efektivitas metode Min-Max dan Z-Score untuk menyesuaikan rentang nilai pada dataset, guna memastikan distribusi data yang seragam dan meningkatkan kinerja model analisis.

Metode Normalisasi Min-Max adalah teknik untuk mengubah data kompleks tanpa menghilangkan informasi yang ada, sehingga data menjadi lebih mudah diolah. Metode ini bekerja dengan menstandarkan data dengan cara menempatkannya dalam rentang 0 hingga 1, di mana nilai terkecil diubah menjadi 0 dan nilai terbesar menjadi 1. Teknik ini membantu untuk menciptakan keseimbangan antara data satu dengan data yang lainnya[21]. Metode Normalisasi Min-Max menggunakan persamaan (2).

$$v' = \frac{v - \min_A}{\max_A - \min_A}(\text{range.max} - \text{range.min}) + \text{range.min} \quad (2)$$

Data yang belum dinormalisasi (v) diubah menjadi data yang sudah dinormalisasi (v'). Proses ini dilakukan dengan menggunakan nilai minimum ($minA$) dan nilai maksimum ($maxA$) dari setiap kolom atau atribut sebelum dinormalisasi. Normalisasi dilakukan dalam rentang nilai tertentu, dalam hal ini dari 0 hingga 1 (range min = 0, range max = 1), sehingga distribusi data menjadi lebih seragam dan dapat mempermudah proses analisis lebih lanjut, seperti dalam metode Backpropagation Neural Network.

Normalisasi Z-score adalah metode normalisasi yang mengubah data berdasarkan nilai rata-rata (mean) dan deviasi standar dari data tersebut. Teknik ini bekerja dengan menghitung z-score untuk setiap data point, yang menunjukkan seberapa jauh data tersebut berada dari mean dalam satuan deviasi standar dengan menggunakan persamaan (3)[22].

$$Z = \frac{(X - \mu)}{\sigma} \quad (3)$$

Data asli (X) sehingga memiliki distribusi dengan rata-rata (μ) nol dan deviasi standar (σ) satu. Proses ini membantu menormalkan variasi dalam skala data, sehingga setiap fitur memiliki kontribusi yang seimbang dalam analisis atau pemodelan. Dengan mengurangi rata-rata dari setiap nilai data asli dan membaginya dengan deviasi standarnya, standarisasi memastikan bahwa data memiliki skala yang seragam, yang penting dalam algoritma pembelajaran mesin.

2.3 Model Klasifikasi

Untuk model klasifikasi, algoritma yang akan digunakan adalah backpropagation. Salah satu algoritma yang sangat sering digunakan dan dikembangkan dalam pengolahan data dari jaringan syaraf tiruan. Berikut adalah rincian algoritma tersebut[23]:

1. Inisialisasikan semua bobot dengan bilangan kecil secara acak
2. Apabila kondisi penghentian masih belum tercapai, maka lanjutkan langkah 2-9
3. Menginisiasi bobot, menentukan learning rate (α), maksimum epoch, dan target error.
   Masuk ke Fase 1 : Propagasi Maju (Forward Propagation)
4. Memasukkan (x_i , i = 1,2,3, … , n) pada tiap unit, lalu menerima dan meneruskan sinyal kepada semua unit pada lapisan selanjutnya.
5. Tiap unit lapisan tersembunyi (z_i, j = 1,2,3, … ,p) menjumlahkan sinyal-sinyal input terbobot dengan rumus (4). Kemudian sinyal tersebut dikirimkan ke semua unit di lapisan berikutnya (lapisan keluaran).

   $$z_{inj} = v_{0j} + \sum_{i=1}^{n} x_1 v_{ji} \quad (4)$$

   Penjumlahan sinyal yang masuk ke neuron pada hidden layer ke-j (z_{inj}) dihitung sebagai akumulasi dari nilai input (x_1) pada input layer ke-i yang dikalikan dengan bobot penghubung antara neuron input layer ke-i dan hidden layer ke-j (v_{ji}). Selain itu, setiap neuron di hidden layer juga memiliki bobot bias (v_{0j}) yang ditambahkan ke total sinyal yang diterima.

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 76−84
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5871

6. Tiap unit pada lapisan keluaran (y_k, k = 1,2,3, … ,m) dijumlahkan bobotnya dengan sinyal. Kemudian hitung nilai keluaran dengan menggunakan fungsi aktivasi.
   Masuk ke Fase II : Propagasi Mundur (Backpropagation)
7. Tiap-tiap unit keluaran (y_k, k = 1,2,3, … ,m) dihitung kesalahan/error dengan rumus (5). Kemudian melakukan perubahan nilai bobot dan nilai bias.

$$\delta_k = (t_k - y_{k)}\, f'\left(y_{in_k}\right) = (t_k - y_k) y_k (1 - y_k) \quad (5)$$

   Faktor koreksi (δ_k) pada neuron output layer ke-k dihitung untuk memperbarui bobot jaringan saraf tiruan. Faktor ini diperoleh dengan menghitung selisih antara target output (t_k) dan output aktual (y_k) pada neuron output layer ke-k. Hasil selisih ini kemudian dikalikan dengan turunan fungsi aktivasi ($f'\left(y_{in_k}\right)$)) terhadap nilai input yang diterima oleh neuron output (y_{in_k}).
8. Tiap-tiap unit hidden layer z_j, (j = 1,2,3, … ,m) dihitung faktor hidden unit berdasarkan kesalahan dengan rumus (6). Kemudian nilai ini dikalikan dengan turunan fungsi aktivasi guna menghitung informasi kesalahan. Selanjutnya, memperbarui nilai bobot dan nilai bias.

$$\delta_{in_j} = \sum_{k=1}^{m} \delta_k w_{jk} \quad (6)$$

   Faktor koreksi (δ_{in_j}) untuk neuron pada hidden layer ke-j dihitung dengan mengakumulasi kontribusi dari faktor koreksi pada neuron output layer ke-k (δ_k) yang dikalikan dengan bobot penghubung antara neuron hidden layer ke-j dan neuron output layer ke-k (w_{jk}).
   Fase III : Perubahan Bobot
9. Tiap-tiap unit output y_(k)(k = 1,2,3, .. , m) dilakukan pembaruan bobot dan biasnya (j = 1,2,3, ... , p)
10. Menguji apakah kondisi penghentian sudah tercapai. Jika belum, maka ulangi Langkah ke 3 sampai 8 hingga mencapai kondisi berhenti

2.4 Evaluasi Model

Tahap evaluasi model merupakan langkah dalam sebuah penelitian untuk menilai kinerja model dan memastikan bahwa model tersebut sesuai dengan tujuan penelitian. Pada tahap ini, akan memberikan gambaran yang lengkap mengenai performa model. Selain itu, tahap evaluasi juga mencakup analisis error untuk memahami jenis kesalahan yang dilakukan model dan potensi penyebabnya. Hasil evaluasi ini kemudian dibandingkan dengan baseline atau target yang telah ditentukan sebelumnya untuk menentukan apakah model memenuhi kriteria keberhasilan. Melalui tahap ini, peneliti dapat mengambil kesimpulan yang informatif mengenai efektivitas dan reliabilitas model yang dikembangkan, serta memberikan dasar untuk rekomendasi atau perbaikan di masa mendatang[24].

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pengumpulan Data

Penulis melakukan identifikasi terhadap masalah yang muncul, yakni perihal krisis air bersih dan layak konsumsi, dengan merujuk pada sejumlah artikel dan jurnal terkait. Pada tahap ini, yang dilakukan adalah mengumpulkan data dengan parameter-parameter yang mempengaruhi kelayakan air minum seperti nilai pH, hardness, solids, dll. Sehingga diperoleh penelitian ini menggunakan data sekunder yang didapatkan dari situs Kaggle yang dapat diakses melalui tautan https://www.kaggle.com/datasets/uom190346a/water-quality-and-potability. Dataset ini berisi pengukuran dan penilaian beberapa parameter yang mempengaruhi kualitas air terkait dengan potability, yaitu kesesuaian air untuk dikonsumsi manusia. Setiap baris dalam kumpulan data mewakili sampel air dengan atribut tertentu, dan kolom "Potabilitas" menunjukkan apakah air tersebut layak konsumsi atau tidak layak konsumsi. Jumlah total data dalam dataset ini mencapai 3276 record, dengan 9 atribut yaitu pH, Hardness, Solids, Chloramines, Sulfate, Conductivity, Organic carbon, Trihalomethanes, dan Turbidity. File dataset yang diunduh dari situs tersebut berformat .csv. Tabel 1 adalah contoh sebagian dataset kelayakan air minum yang akan digunakan dalam penelitian ini:

Tabel 1. Dataset yang digunakan

No.	ph	Hardness	Solids	Chloramines	Sulfate	Conductivity	Organic carbon	Trihalomethanes	Turbidity	Potability
1		204,890	20791,319	7,300	368,516	564,309	10,380	86,991	2,963	0
2	3,716	129,423	18630,058	6,635		592,885	15,180	56,329	4,501	0
3	8,099	224,236	19909,542	9,276		418,606	16,869	66,420	3,056	0
4	8,317	214,373	22018,417	8,059	356,886	363,267	18,437	100,342	4,629	0
5	9,092	181,102	17978,986	6,547	310,136	398,411	11,558	31,998	4,075	0
6	5,584	188,313	28748,688	7,545	326,678	280,468	8,400	54,918	2,560	0
7	10,224	248,072	28749,717	7,513	393,663	283,652	13,790	84,604	2,673	0

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 76−84
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5871

No.	ph	Hardness	Solids	Chloramines	Sulfate	Conductivity	Organic carbon	Trihalomethanes	Turbidity	Potability
8	8,636	203,362	13672,092	4,563	303,310	474,608	12,364	62,798	4,401	0
9		118,989	14285,584	7,804	268,647	389,376	12,706	53,929	3,595	0
10	11,180	227,231	25484,508	9,077	404,042	563,885	17,928	71,977	4,371	0
11	7,361	165,521	32452,614	7,551	326,624	425,383	15,587	78,740	3,662	0
12	7,975	218,693	18767,657	8,110		364,098	14,526	76,486	4,012	0
13	7,120	156,705	18730,814	3,606	282,344	347,715	15,930	79,501	3,446	0
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
3276	7,875	195,102	17404,177	7,509		327,460	16,140	78,698	2,309	1

3.2 Hasil Pemrosesan Data

Dataset yang digunakan masih memiliki missing value. Mengidentifikasi missing value dilakukan dengan tujuan untuk melihat seberapa banyak kesalahan dalam data. Di bawah ini terdapat gambar yang menunjukkan distribusi missing value dalam dataset:

```
# Menampilkan jumlah missing value untuk setiap kolom
print(datawater.isnull().sum())
```

```
ph                491
Hardness            0
Solids              0
Chloramines         0
Sulfate           781
Conductivity        0
Organic_carbon      0
Trihalomethanes   162
Turbidity           0
Potability          0
dtype: int64
```

Gambar 2. Distribusi Missing Value

Pada gambar 2, dapat diamati distribusi missing value pada berbagai fitur dalam dataset. Secara spesifik, fitur pH memiliki 491 missing value, yang menunjukkan adanya kesenjangan data yang signifikan. Demikian pula, fitur Sulfate mengalami 781 missing value, yang merupakan jumlah terbesar di antara fitur lainnya. Selain itu, fitur trihalomethanes juga menunjukkan adanya 162 missing value. Informasi ini sangat penting untuk langkah selanjutnya dalam proses pembersihan data, di mana strategi seperti imputasi atau penggantian nilai pada baris yang mengandung missing value dapat diterapkan untuk memastikan bahwa analisis data tetap valid dan dapat diandalkan. Teknik imputasi ini digunakan agar tetap mempertahankan jumlah utuh dari dataset semula. Berikut adalah data hasil imputasi missing value menggunakan nilai rata-rata:

Tabel 2. Dataset hasil penggantian missing value dengan nilai rata-rata

No.	ph	Hardness	Solids	Chloramines	Sulfate	Conductivity	Organic carbon	Trihalomethanes	Turbidity	Potability
1	7,081	204,890	20791,319	7,300	368,516	564,309	10,380	86,991	2,963	0
2	3,716	129,423	18630,058	6,635	333,776	592,885	15,180	56,329	4,501	0
3	8,099	224,236	19909,542	9,276	333,776	418,606	16,869	66,420	3,056	0
4	8,317	214,373	22018,417	8,059	356,886	363,267	18,437	100,342	4,629	0
5	9,092	181,102	17978,986	6,547	310,136	398,411	11,558	31,998	4,075	0
6	5,584	188,313	28748,688	7,545	326,678	280,468	8,400	54,918	2,560	0
7	10,224	248,072	28749,717	7,513	393,663	283,652	13,790	84,604	2,673	0
8	8,636	203,362	13672,092	4,563	303,310	474,608	12,364	62,798	4,401	0
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
3276	7,875	195,102	17404,177	7,509	333,776	327,460	16,140	78,698	2,309	1

Selanjutnya sebagai pembanding, proses pembersihan data juga dilakukan dengan metode penghapusan data yang memiliki missing value. Berikut adalah data hasil penghapusan missing value:

Tabel 3. Dataset hasil penghapusan missing value

No.	ph	Hardness	Solids	Chloramines	Sulfate	Conductivity	Organic carbon	Trihalomethanes	Turbidity	Potability
1	8,317	214,373	22018,417	8,059	356,886	363,267	18,437	100,342	4,629	0
2	9,092	181,102	17978,986	6,547	310,136	398,411	11,558	31,998	4,075	0
3	5,584	188,313	28748,688	7,545	326,678	280,468	8,400	54,918	2,560	0

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 76−84
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5871

No.	ph	Hardness	Solids	Chloramines	Sulfate	Conductivity	Organic carbon	Trihalomethanes	Turbidity	Potability
4	10,224	248,072	28749,717	7,513	393,663	283,652	13,790	84,604	2,673	0
5	8,636	203,362	13672,092	4,563	303,310	474,608	12,364	62,798	4,401	0
6	11,180	227,231	25484,508	9,077	404,042	563,885	17,928	71,977	4,371	0
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
2011	4,668	193,682	47580,992	7,167	359,949	526,424	13,894	66,688	4,436	1

Sebagai hasil penghapusan dari langkah ini, jumlah record yang awalnya mencapai 3.276 berkurang secara substansial menjadi 2.011 record setelah proses pembersihan dilakukan. Dengan demikian, dataset yang telah diperbaiki ini diharapkan dapat memberikan hasil analisis yang lebih akurat dan valid tanpa adanya distorsi akibat data yang tidak lengkap.

Tahap berikutnya dalam pengolahan data adalah melakukan normalisasi untuk memastikan bahwa semua fitur dalam dataset memiliki skala yang konsisten. Normalisasi yang akan digunakan dalam langkah ini adalah metode MinMax Scaler. Metode MinMax Scaler bekerja dengan mengubah setiap nilai data sehingga berada dalam rentang nilai tertentu, yaitu dari 0 hingga 1. Proses ini dilakukan dengan mengurangkan nilai minimum dari setiap fitur dan kemudian membaginya dengan rentang (perbedaan antara nilai maksimum dan minimum) dari fitur tersebut. Tabel 4 menunjukkan hasil normalisasi menggunakan MinMax Scaler pada dataset di tabel 2.

Tabel 4. Dataset hasil normalisasi MinMax Scaler

No.	ph	Hardness	Solids	Chloramines	Sulfate	Conductivity	Organic carbon	Trihalomethanes	Turbidity	Potability
1	0,506	0,571	0,336	0,544	0,680	0,669	0,313	0,700	0,286	0
2	0,265	0,297	0,301	0,492	0,582	0,719	0,497	0,451	0,577	0
3	0,579	0,641	0,322	0,699	0,582	0,415	0,562	0,533	0,304	0
4	0,594	0,606	0,356	0,603	0,647	0,318	0,622	0,808	0,601	0
5	0,649	0,485	0,290	0,485	0,515	0,379	0,359	0,254	0,496	0
6	0,399	0,511	0,467	0,563	0,562	0,173	0,238	0,440	0,210	0
7	0,730	0,728	0,467	0,561	0,752	0,179	0,444	0,680	0,231	0
8	0,617	0,566	0,219	0,330	0,495	0,513	0,389	0,503	0,558	0
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
3276	0,562	0,536	0,280	0,560	0,582	0,255	0,534	0,632	0,162	1

Normalisasi yang digunakan dalam langkah ini adalah metode z-score. Metode z-score merupakan teknik normalisasi yang mengubah data sehingga setiap fitur memiliki rata-rata nol dan deviasi standar satu. Hal ini dilakukan dengan menghitung selisih antara setiap nilai data dengan rata-rata dari seluruh data dalam fitur tersebut, kemudian membaginya dengan deviasi standar. Setelah data dinormalisasi menggunakan z-score, setiap nilai data akan berada pada rentang yang lebih homogen, biasanya berkisar antara -3 hingga 3, yang merepresentasikan jarak nilai tersebut dari rata-rata dalam satuan deviasi standar. Tabel 5 menunjukkan hasil normalisasi dengan z-score pada dataset di tabel 2.

Tabel 5. Dataset hasil normalisasi Z-score

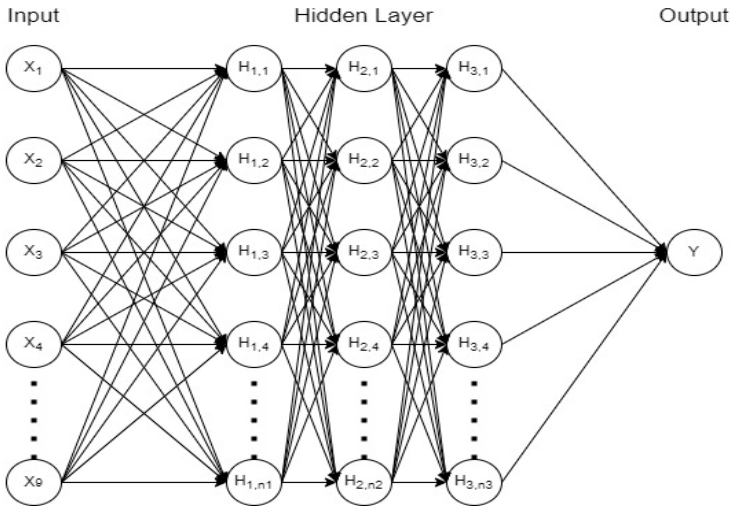
No.	ph	Hardness	Solids	Chloramines	Sulfate	Conductivity	Organic Carbon	Trihalomethanes	Turbidity	Potability
1	0,782	0,564	0,012	0,584	0,574	-0,784	1,227	2,112	0,845	0
2	1,275	-0,456	-0,456	-0,371	-0,560	-0,348	-0,842	-2,140	0,135	0
3	-0,955	-0,235	0,791	0,259	-0,159	-1,810	-1,792	-0,714	-1,807	0
4	1,995	1,597	0,791	0,239	1,467	-1,771	-0,171	1,132	-1,662	0
5	0,985	0,227	-0,954	-1,623	-0,726	0,596	-0,600	-0,224	0,553	0
6	2,603	0,958	0,413	1,226	1,719	1,702	1,074	0,347	0,514	0
7	0,175	-0,933	1,219	0,263	-0,160	-0,014	0,370	0,768	-0,394	0
8	0,022	-1,203	-0,369	-2,227	-1,235	-0,977	0,473	0,815	-0,672	0
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
3276	-1,537	-0,070	2,970	0,020	0,649	1,238	-0,139	0,018	0,597	1

3.3 Hasil Model Klasifikasi

Dalam penelitian ini, model klasifikasi menggunakan Backpropagation Neural Network (BPNN) diterapkan pada empat variasi dataset yang dihasilkan dari cara penanganan missing value dan normalisasi. Dua pendekatan digunakan untuk menangani missing value: penghapusan seluruh baris yang mengandung nilai kosong dan penggantian nilai kosong dengan nilai rata-rata. Setelah itu, kedua jenis dataset ini dinormalisasi menggunakan dua metode: MinMax Scaler dan Z-score. Untuk memastikan bahwa model klasifikasi Backpropagation Neural Network (BPNN) yang dikembangkan memiliki kinerja yang stabil dan dapat diandalkan, maka digunakan teknik pembagian data yang

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 76−84
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5871

dikenal sebagai K-fold cross-validation. Secara khusus melakukan eksperimen dengan menggunakan K-fold dengan nilai K=10 dan K=7. Pada tahap pelatihan model klasifikasi menggunakan Backpropagation Neural Network (BPNN), iterasi maksimal ditetapkan sebanyak 2000 untuk memastikan proses pembaruan bobot dapat berlangsung hingga model mencapai konvergensi atau mendekati performa optimal. Dalam proses ini, dua fungsi aktivasi utama: tanh dan ReLU dibandingkan untuk menentukan mana yang memberikan hasil terbaik.



Gambar 3. Arsitektur Jaringan Syaraf Tiruan

Model akan menerapkan arsitektur jaringan syaraf tiruan seperti pada gambar 3. Untuk konfigurasi hidden layer, model pertama kali diuji dengan hanya satu hidden layer yang berisi 9 hingga 18 neuron untuk menentukan jumlah neuron yang memberikan akurasi terbaik. Setelah menemukan konfigurasi terbaik pada satu layer, model kemudian dikembangkan menjadi dua hidden layer. Pada konfigurasi dua layer ini, layer pertama menggunakan jumlah neuron yang telah menunjukkan akurasi terbaik dari tahap sebelumnya, sedangkan layer kedua divariasikan jumlah neuronnya antara 9 hingga 18 untuk mencari kombinasi yang paling optimal. Percobaan serupa dilakukan dengan menambahkan hidden layer ketiga, di mana layer pertama dan kedua tetap dengan jumlah neuron yang ditemukan paling optimal dari uji sebelumnya, sementara layer ketiga divariasikan dengan jumlah neuron yang sama, yakni 9 hingga 18. Proses iteratif ini bertujuan untuk menemukan arsitektur yang paling efektif dalam meningkatkan akurasi model dalam klasifikasi data. Maka, berdasarkan skenario percobaan di atas, hasil terbaik yang didapatkan setiap skenario adalah sebagai berikut:

a. Menggunakan K-fold = 10 dan aktivasi ReLU

Tabel 6. Hasil model klasifikasi menggunakan K-fold = 10 dan aktivasi ReLU

Missing Value	Normalisasi	Hidden Layer	Akurasi	Presisi	Recall	F-1 Score
Menghapus Missing Value	MinMax Scaler	[9 ; 18 ; 9]	0,649882	0,598121	0,341444	0,401164
	Z-Score	[10]	0,644948	0,613917	0,410263	0,471379
Mengganti Missing Value Dengan Nilai Rata-Rata	MinMax Scaler	[16 ; 18]	0,632528	0,576429	0,299229	0,370429
	Z-Score	[9]	0,619336	0,54666	0,320598	0,380382

Pada Tabel 6, dapat dilihat hasil skenario percobaan yang optimal adalah akurasi 0,649882 pada data yang dilakukan proses penghapusan missing value, normalisasi MinMax Scaler, dan 9 neuron pada hidden layer pertama, 18 neuron pada hidden layer kedua, dan 9 neuron pada hidden layer ketiga.

b. Menggunakan K-fold = 10 dan aktivasi tanh

Tabel 7. Hasil model klasifikasi menggunakan K-fold = 10 dan aktivasi tanh

Missing Value	Normalisasi	Hidden Layer	Akurasi	Presisi	Recall	F-1 Score
Menghapus Missing Value	MinMax Scaler	[15 ; 9 ; 15]	0,627479	0,39769	0,092862	0,130319
	Z-Score	[12]	0,645429	0,597591	0,443732	0,489807
Mengganti Missing Value Dengan Nilai Rata-Rata	MinMax Scaler	[9 ; 13 ; 14]	0,649211	0,535754	0,212568	0,294848
	Z-Score	[10]	0,617496	0,545248	0,33882	0,394003

Pada Tabel 7, dapat dilihat hasil skenario percobaan yang optimal adalah akurasi 0,649211 pada data yang dilakukan proses penggantian missing value dengan nilai rata-rata, normalisasi MinMax Scaler, dan 9 neuron pada hidden layer pertama, 13 neuron pada hidden layer kedua, dan 14 neuron pada hidden layer ketiga.

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 76−84
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5871

c. Menggunakan K-fold = 7 dan aktivasi ReLU

Tabel 8. Hasil model klasifikasi menggunakan K-fold = 7 dan aktivasi ReLU

Missing Value	Normalisasi	Hidden Layer	Akurasi	Presisi	Recall	F-1 Score
Menghapus Missing Value	MinMax Scaler	[14 ; 14]	0,653403	0,534602	0,341779	0,412533
	Z-Score	[10]	0,642983	0,579957	0,406535	0,471984
Mengganti Missing Value Dengan Nilai Rata-Rata	MinMax Scaler	[17 ; 17]	0,641331	0,597149	0,304932	0,389554
	Z-Score	[13]	0,637057	0,557883	0,340223	0,413184

Pada Tabel 7, dapat dilihat hasil skenario percobaan yang optimal adalah akurasi 0,653403 pada data yang dilakukan proses penghapusan missing value, normalisasi MinMax Scaler, dan 14 neuron pada hidden layer pertama dan 14 neuron pada hidden layer kedua.

d. Menggunakan K-fold = 7 dan aktivasi tanh

Tabel 9. Hasil model klasifikasi menggunakan K-fold = 7 dan aktivasi tanh

Missing Value	Normalisasi	Hidden Layer	Akurasi	Presisi	Recall	F-1 Score
Menghapus Missing Value	MinMax Scaler	[17; 15; 10]	0,657873	0,551743	0,32433	0,402757
	Z-Score	[9]	0,63453	0,562762	0,413584	0,471524
Mengganti Missing Value Dengan Nilai Rata-Rata	MinMax Scaler	[10; 12; 11]	0,634921	0,373383	0,148952	0,204944
	Z-Score	[9]	0,635836	0,568697	0,319724	0,396033

Pada Tabel 7, dapat dilihat hasil skenario percobaan yang optimal adalah akurasi 0,657873 pada data yang dilakukan proses penghapusan missing value, normalisasi MinMax Scaler, dan 17 neuron pada hidden layer pertama, 15 neuron pada hidden layer kedua, dan 10 neuron pada hidden layer ketiga.

3.4 Hasil Evaluasi Model

Hasil evaluasi model pada penelitian ini mengungkapkan bahwa untuk dataset yang digunakan, penghapusan missing value lebih efektif dibandingkan dengan imputasi menggunakan nilai rata-rata, karena metode penghapusan cenderung memberikan data yang lebih bersih dan mencegah pengenalan bias yang mungkin terjadi dari nilai rata-rata. Dalam hal normalisasi, MinMax Scaler terbukti lebih efektif daripada Z-score normalization. Hal ini ditunjukkan oleh pola yang muncul dalam penelitian, di mana normalisasi dengan Z-score menghasilkan akurasi yang relatif lebih tinggi ketika model menggunakan satu lapis hidden layer. Sebaliknya, ketika menggunakan normalisasi MinMax Scaler, model menunjukkan performa yang lebih baik dengan struktur yang lebih kompleks, yakni dengan dua atau tiga lapis hidden layer. Temuan ini menunjukkan bahwa MinMax Scaler mampu menjaga hubungan proporsional antar fitur lebih baik dalam arsitektur model yang lebih dalam, sementara Z-score normalization lebih sesuai untuk model dengan struktur yang lebih sederhana. Dengan demikian, pemilihan teknik pengolahan data dan arsitektur model yang sesuai sangat penting untuk memaksimalkan kinerja klasifikasi pada dataset ini.

Pada penelitian ini, hasil evaluasi model menunjukkan bahwa model terbaik adalah dengan menggunakan dataset yang telah melalui pemrosesan penghapusan missing value dan normalisasi dengan MinMax Scaler, serta menggunakan arsitektur dengan hidden layer [17;15;10], menghasilkan akurasi sebesar 0,6579 atau 65%. Meskipun ini adalah model terbaik yang diperoleh, akurasi ini masih dianggap kurang memadai untuk diaplikasikan.

Beberapa alasan yang mungkin menyebabkan akurasi yang belum optimal antara lain adalah kemungkinan distribusi data yang tidak seimbang dan juga mungkin metode pemrosesan data yang diterapkan masih terbatas pada MinMax Scaler dan Z-score.

4. KESIMPULAN

Dalam penelitian ini, metode Backpropagation Neural Network (BPNN) telah berhasil diterapkan untuk melakukan klasifikasi kelayakan air minum. Hasil evaluasi menunjukkan bahwa model yang dihasilkan mampu mencapai tingkat akurasi sebesar 0,6579. Meskipun begitu akurasi ini masih perlu ditingkatkan lagi demi mencapai hasil yang lebih baik dan memuaskan. Untuk penelitian selanjutnya, disarankan untuk mempertimbangkan beberapa aspek. Pertama, perlu dilakukan eksplorasi lebih lanjut terhadap metode pengolahan data lainnya, seperti teknik normalisasi yang lebih canggih atau metode penanganan missing value yang lebih adaptif, seperti imputasi berbasis model. Kedua, disarankan untuk mencoba arsitektur neural network yang lebih kompleks atau menggunakan kombinasi metode machine learning lainnya untuk meningkatkan kinerja model. Dengan memperhatikan saran-saran ini, diharapkan penelitian selanjutnya akan memberikan hasil yang lebih memuaskan dalam klasifikasi dan analisis dataset yang serupa.

REFERENCES

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 76−84
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5871

[1] WHO, "Drinking-water," who.int. Accessed: Dec. 17, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/drinking-water
[2] H. Istiawan, "Hari Air Sedunia, Kondisi Sungai di Indonesia Memprihatinkan," Okezone.com. [Online]. Available: https://nasional.okezone.com/read/2017/03/22/337/1648855/hari-air-sedunia-kondisi-sungai-di-indonesia-memprihatinkan
[3] Tumangger and R. M. Sahalatua, "Komparasi Metode Data Mining Support Vector Machine Dengan Naive Bayes Untuk Klasifikasi Status Kualitas Air," Universitas Brawijaya, 2020.
[4] A. Safira, L. M. Sarudi As., A. Puspitasari, N. M. E. Normasari, and A. P. Rifai, "Pengembangan Neural Network Untuk Prediksi Kualitas Air," J. Rekavasi, vol. 10, no. 2, pp. 30–36, 2023, doi: 10.34151/rekavasi.v10i2.4014.
[5] A. Herdiansah, R. I. Borman, D. Nurnaningsih, A. A. J. Sinlae, and R. R. Al Hakim, "Klasifikasi Citra Daun Herbal Dengan Menggunakan Backpropagation Neural Networks Berdasarkan Ekstraksi Ciri Bentuk," JURIKOM (Jurnal Ris. Komputer), vol. 9, no. 2, p. 388, 2022, doi: 10.30865/jurikom.v9i2.4066.
[6] C. Sekhar and P. S. Meghana, "A Study on Backpropagation in Artificial Neural Networks," Asia-Pacific J. Neural Networks Its Appl., vol. 4, no. 1, pp. 21–28, 2020, doi: 10.21742/ajnnia.2020.4.1.03.
[7] N. Norhikmah and R. Rumini, "Klasifikasi Peminjaman Buku Menggunakan Neural Network Backpropagation," Sistemasi, vol. 9, no. 1, p. 1, 2020, doi: 10.32520/stmsi.v9i1.562.
[8] I. K. P. Suniantara, G. Suwardika, and S. Soraya, "Peningkatan Akurasi Klasifikasi Ketidaktepatan Waktu Kelulusan Mahasiswa Menggunakan Metode Boosting Neural Network," J. Varian, vol. 3, no. 2, pp. 95–102, 2020, doi: 10.30812/varian.v3i2.651.
[9] F. Izhari, M. Zarlis, and Sutarman, "Analysis of backpropagation neural neural network algorithm on student ability based cognitive aspects," IOP Conf. Ser. Mater. Sci. Eng., vol. 725, no. 1, pp. 243–252, 2020, doi: 10.1088/1757-899X/725/1/012103.
[10] A. Yuberta, "Jaringan Syaraf Tiruan dengan Algoritma Backpropagation dalam Memprediksi Hasil Asesmen Nasional Berbasis Komputer (ANBK) SMP Se Kota Sawahlunto," J. Inf. dan Teknol., vol. 4, no. 4, pp. 200–205, 2022, doi: 10.37034/jidt.v4i4.234.
[11] K. Anwar and Manuharawati, "Implementasi Metode Backpropagation Neural Network Dalam Meramalkan Tingkat Inflasi Di Indonesia," J. Ilm. Mat., vol. 9, no. 2, pp. 437–446, 2021, [Online]. Available: https://media.neliti.com/media/publications/249234-model-infeksi-hiv-dengan-pengaruh-percob-b7e3cd43.pdf
[12] L. Savitri and R. Nursalim, "Klasifikasi Kualitas Air Minum menggunakan Penerapan Algoritma Machine Learning dengan Pendekatan Supervised Learning," Diophantine J. Math. Its Appl., vol. 2, no. 01, pp. 30–36, 2023, doi: 10.33369/diophantine.v2i01.28260.
[13] A. Tangkelayuk, "The Klasifikasi Kualitas Air Menggunakan Metode KNN, Naïve Bayes, dan Decision Tree," JATISI (Jurnal Tek. Inform. dan Sist. Informasi), vol. 9, no. 2, pp. 1109–1119, 2022, doi: 10.35957/jatisi.v9i2.2048.
[14] M. Nazar Yuniar, "Klasifikasi Kualitas Air Bersih Menggunakan Metode Naïve baiyes," J. Sains dan Teknol., vol. 5, no. 1, pp. 243–246, 2023, [Online]. Available: https://doi.org/10.55338/saintek.v5i1.1383
[15] M. M. Mutoffar and A. Fadillah, "Klasifikasi Kualitas Air Sumur Menggunakan Algoritma Random Forest," Naratif J. Nas. Riset, Apl. dan Tek. Inform., vol. 4, no. 2, pp. 138–146, 2022, doi: 10.53580/naratif.v4i2.160.
[16] Y. V. Sari, Z. Muallifah, and A. Fanani, "Klasifikasi Kualitas Air Menggunakan Metode Extreme Learning Machine (ELM)," J. JUPITER, vol. 15, no. 2, pp. 983–994, 2023, [Online]. Available: https://jurnal.polsri.ac.id/index.php/jupiter/article/view/6995
[17] A. Rizky Fadilla and P. Ayu Wulandari, "Literature Review Analisis Data Kualitatif: Tahap PengumpulanData," Mitita J. Penelit., vol. 1, no. No 3, pp. 34–46, 2023.
[18] Ardiansyah, Risnita, and M. S. Jailani, "Teknik Pengumpulan Data Dan Instrumen Penelitian Ilmiah Pendidikan Pada Pendekatan Kualitatif dan Kuantitatif," J. IHSAN J. Pendidik. Islam, vol. 1, no. 2, pp. 1–9, 2023, doi: 10.61104/ihsan.v1i2.57.
[19] A. Desiani, N. R. Dewi, A. N. Fauza, N. Rachmatullah, M. Arhami, and M. Nawawi, "Handling Missing Data Using Combination of Deletion Technique, Mean, Mode and Artificial Neural Network Imputation for Heart Disease Dataset," Sci. Technol. Indones., vol. 6, no. 4, pp. 303–312, 2021, doi: 10.26554/sti.2021.6.4.303-312.
[20] W. Sudrajat and I. Cholid, "K-Nearest Neighbor (K-Nn) Untuk Penanganan Missing Value Pada Data Umkm," J. Rekayasa Sist. Inf. dan Teknol., vol. 1, no. 2, pp. 54–63, 2023, doi: 10.59407/jrsit.v1i2.77.
[21] E. Patimah, V. B. Haekal, and D. Sandya Prasvita, "Klasifikasi Penyakit Liver dengan Menggunakan Metode Decision Tree," Semin. Nas. Mhs. Ilmu Komput. dan Apl. Jakarta-Indonesia, vol. 2, no. 1, pp. 655–659, 2021.
[22] R. G. Whendasmoro and J. Joseph, "Analisis Penerapan Normalisasi Data Dengan Menggunakan Z-Score Pada Kinerja Algoritma K-NN," JURIKOM (Jurnal Ris. Komputer), vol. 9, no. 4, p. 872, 2022, doi: 10.30865/jurikom.v9i4.4526.
[23] H. Putra and N. Ulfa Walmi, "Penerapan Prediksi Produksi Padi Menggunakan Artificial Neural Network Algoritma Backpropagation," J. Nas. Teknol. dan Sist. Inf., vol. 6, no. 2, pp. 100–107, 2020, doi: 10.25077/teknosi.v6i2.2020.100-107.
[24] M. Shofwan Khamid et al., "Prediksi Jumlah Sampah Kelurahan Menggunakan Neural Network Backpropagation," J. Inf. Syst. Res., vol. 5, no. 2, pp. 713–721, 2024, doi: 10.47065/josh.v5i2.4825.