

KLASIFIKASI KUALITAS AIR MENGGUNAKAN METODE EXTREME LEARNING MACHINE (ELM)

Yana Vita Sari¹, Zumrotul Muallifah², Aris Fanani³

^{1,2,3} Program Studi Matematika, Fakultas Sains dan Teknologi, UIN Sunan Ampel Surabaya
Gunung Anyar, Surabaya, East Java 60294

e-mail: *yanavitasari05@gmail.com, zumrotulmuallifah@gmail.com,
arisfanani@uinsby.ac.id

Abstrak

Air merupakan senyawa yang penting bagi semua makhluk hidup di bumi. Kebutuhan air bersih meningkat seiring dengan berjalannya waktu, hal tersebut berbanding terbalik dengan ketersediannya di alam. Hal ini disebabkan karena banyaknya pembangunan tanpa memperhatikan keseimbangan lingkungan sekitar sehingga semakin sedikitnya daerah resapan air terutama di daerah perkotaan. Hal tersebut dapat berakibat dengan penurunan kualitas air yang menyebabkan urgensi kebutuhan untuk mengawasi, menilai, dan mengklasifikasi kualitas air yang layak untuk dikonsumsi. Pemeriksaan kualitas dari air merupakan salah satu upaya untuk mengontrol ada tidaknya penyakit, bakteri maupun kuman dalam air sehingga dapat mengelompokkan air yang layak diminum dan tidak. Tujuan dari penelitian ini adalah untuk mengklasifikasi kualitas air berdasarkan parameter meliputi zat dalam air yang diperoleh dari sumber air tawar dengan menggunakan metode ELM untuk mengklasifikasi air yang dapat diminum atau tidak. Data yang digunakan pada penelitian ini diperoleh dari web kaggle. Dari jumlah data yang dimiliki pada kedua kelas tidak seimbang atau dapat disebut dengan imbalance data sehingga perlu diatasi dengan menggunakan metode SMOTE. Pada penelitian ini melakukan uji coba perbandingan data parameter node hidden. Berdasarkan uji coba diperoleh model optimum pada k-fold 10 dan node hidden 100 menghasilkan akurasi 97.873%, sensitifitas 96.4789%, dan spesifisitas 98.6408%. Hal tersebut membuktikan bahwa metode SMOTE berpengaruh terhadap hasil uji coba.

Kata kunci— ELM, Klasifikasi, Kualitas Air, Pencemaran Air.

Abstract

Water is an important compound for all living things on earth. The need for clean water increases over time, this is inversely proportional to its availability in nature. This is due to the large amount of development without paying attention to the balance of the surrounding environment, resulting in fewer and fewer water catchment areas, especially in urban areas. This can result in a decrease in water quality which creates an urgent need to monitor, assess and classify water quality that is suitable for consumption. Checking the quality of water is an effort to control the presence or absence of disease, bacteria, and germs in the water so that water can be classified as suitable for drinking or not. The aim of this research is to classify water quality based on parameters including substances in water obtained from fresh water sources using the ELM method to classify water as drinkable or not. The data used in this research was obtained from the Kaggle web. The amount of data held in the two classes is not balanced or can be called imbalance data, so it needs to be overcome using the SMOTE method. In this research, we carried out comparison trials of hidden node parameter data. Based on trials, it was obtained that the optimum model at k-fold 10 and hidden nodes 100 produced an accuracy of 97.873%, sensitivity of 96.4789%, and specificity of 98.6408%. This proves that the SMOTE method influences the test results.

Keywords— Classification, ELM, Water Quality, Water Pollution

1. PENDAHULUAN

Air merupakan senyawa yang penting bagi semua makhluk hidup di bumi. Sebesar 70% planet bumi terdiri atas air yang dapat ditemukan di laut, lapisan-lapisan es, sungai, permukaan air tawar, danau, uap air, dan lain sebagainya [1]. Air dalam objek-objek tersebut berpindah mengikuti suatu siklus air yaitu penguapan, hujan, dan aliran air di atas permukaan tanah menuju laut [2]. Air dapat berwujud padatan (es), cairan (air) dan gas (uap air). Air memiliki peran dan fungsi biologis, ekologis, sosial ekonomi dan sumberdaya untuk keberlangsungan kehidupan [3].

Kebutuhan air bersih meningkat seiring dengan berjalannya waktu, hal tersebut berbanding terbalik dengan ketersediannya di alam [4]. Hal ini disebabkan karena banyaknya pembangunan tanpa memperhatikan keseimbangan lingkungan sekitar sehingga semakin sedikitnya daerah resapan air terutama di daerah perkotaan. Selain itu, penyebab dari kurangnya ketersediaan air yaitu pengambilan air secara berlebihan, tinggaknya pencemaran terhadap sumber-sumber air, adanya konflik kepentingan ekonomi yang didukung oleh kebijakan yang kurang tepat. Hal tersebut dapat berakibat dengan penurunan kualitas air yang menyebabkan urgensi kebutuhan untuk mengawasi, menilai, dan mengklasifikasi kualitas air yang layak untuk dikonsumsi. Standar kualitas air yang layak diminum yaitu jernih, tidak berbau, tidak berasa, bersuhu wajar, bersih dari bakteri dan kuman serta mengandung sedikit jumlah mineral.

Pemeriksaan kualitas dari air merupakan salah satu upaya untuk mengontrol ada tidaknya penyakit, bakteri maupun kuman dalam air sehingga dapat mengelompokkan air yang layak diminum dan tidak. Seperti pada penelitian oleh Hardiana Said dkk menggunakan algoritma K-Nearest Neighbor untuk memprediksi kualitas air dengan tingkat akurasi tertinggi sebesar 85,24% [5]. Penelitian oleh Ibrahim A. Sukamari dkk juga menunjukkan bahwa indeks kualitas air dapat digunakan untuk menilai potabilitas air [6]. Selain itu, penelitian oleh S Mohd-Asharuddin dkk pada Karakteristik Kualitas Air Bendungan Sembrong, Johor, Malaysia menunjukkan bahwa air Bendungan Sembrong tidak tercemar logam berat karena pembacaan logam berat yang tercatat berada jauh di bawah Kelas I [7].

Pengukuran dan penentuan kualitas air dapat menggunakan metode manual seperti Indeks Pencemaran (IP), Water Quality Index (WQI) dan STORET tetapi metode tersebut memiliki kendala dalam waktu dan biaya yang cukup tinggi sehingga diperlukannya metode lain salah satunya dengan Extreme Learning Machine (ELM) [8]. ELM merupakan jaringan syaraf tiruan feed-forward dengan satu hidden layer [9]. ELM memiliki fitur yang menarik dan signifikan. Variabel-variabel dalam penentuan kualitas air dapat dijadikan pedoman atau acuan dalam membantu mengklasifikasikan kualitas air. Kelebihan dari ELM yaitu dalam learning speed dan tingkat akurasi yang lebih baik dibandingkan dengan metode lainnya. Seperti pada penelitian yang dilakukan oleh Henny dan Kusno yang membandingkan metode ELM dan Backpropagation untuk mengetahui prestasi akademik menghasilkan ELM mempunyai tingkat akurasi tinggi dan kesalahan yang rendah sebesar 14,8% dibandingkan backpropagation sebesar 28,20% [10].

Berdasarkan latar belakang diatas, maka tujuan dari penelitian ini adalah untuk mengklasifikasi kualitas air berdasarkan parameter meliputi zat dalam air yang diperoleh dari sumber air tawar dengan menggunakan metode ELM untuk mengklasifikasi air yang dapat diminum atau tidak. Dengan demikian, penelitian ini diharapkan dapat membantu perencanaan strategi dalam menjaga kualitas, kuantitas, dan kontinuitas air bersih.

2. METODE PENELITIAN

Penelitian ini melibatkan beberapa tahapan. Tahap pertama adalah mengumpulkan data yang akan digunakan dalam proses pelatihan dan pengujian. Selanjutnya, dilakukan perancangan arsitektur klasifikasi kualitas air. Setelah itu, penelitian dilanjutkan dengan tahap implementasi arsitektur sistem yang telah dirancang. Tahap terakhir adalah pengujian yang

bertujuan untuk mengevaluasi kinerja analisis sentimen menggunakan algoritma Extreme Learning Machine.

2.1 Pengumpulan Data

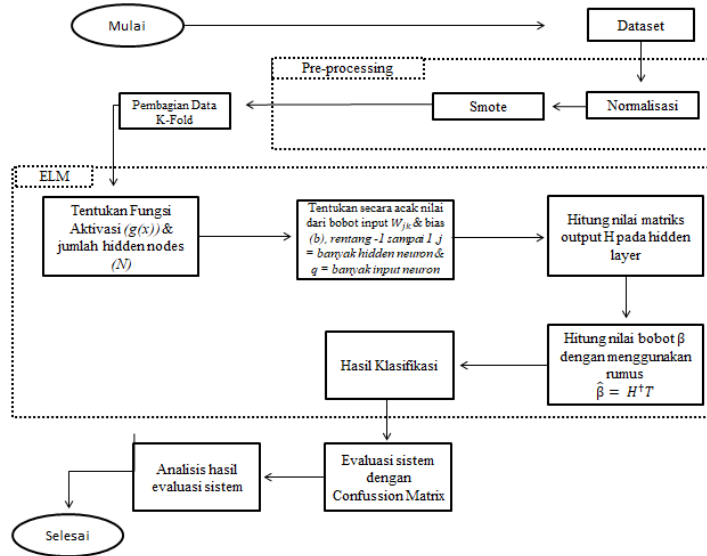
Dalam penelitian ini menggunakan data kuantitatif yang berisi mengenai parameter yang berpengaruh terhadap kualitas air. Data yang diperoleh sebanyak 20 parameter dengan 8000 total data dimana parameter air tersebut yaitu Aluminium, Ammonia, Arsenic, Barium, Cadmium, Chloramine, Copper, Flouride, Bacteria, Viruses, Lead, Nitrates, Nitrites, Mercury, Perchlorate, Radium, Selenium, Silver, Uranium, dan Potabilitas. Data tersebut diambil dari kaggle pada 18 April 2023.

Tabel 1 Data Kualitas Air

No	Aluminium	Amonia	...	Potabilitas
1	1.65	9.08	...	1
2	2.32	21.16	...	1
3	1.01	14.02	⋮	0
4	1.36	11.33	...	1
⋮	⋮	⋮	⋮	⋮
8000	0.04	6.85	...	1

2.2 Alur Penelitian

Secara umum diagram alir penelitian ditunjukkan pada Gambar 1 merupakan proses yang dilakukan untuk mengklasifikasi kualitas air menggunakan metode Extreme Learning Machine (ELM). Langkah pertama pada penelitian ini yaitu input data kualitas air yang diperoleh dari web kaggle yang terdiri atas 2 kelas yaitu air yang layak minum sebanyak 652 dan air yang tidak layak minum sebanyak 348. Dari data tersebut memiliki rentan yang berbeda sehingga perlu dilakukan normalisasi data agar data tetap konstan. Dari jumlah data yang dimiliki pada kedua kelas tidak seimbang atau dapat disebut dengan imbalance data sehingga perlu diatasi dengan menggunakan metode SMOTE. Selanjutnya dilakukan pembagian data training dan data testing menggunakan metode k-fold. Lalu dilanjutkan dengan proses klasifikasi menggunakan metode Extreme Learning Machine Learning. Agar diperoleh model yang optimum, pada penelitian ini melakukan uji coba perbandingan data menggunakan k-fold dengan nilai k 5 dan 10 serta parameter node hidden yaitu 5, 25, 50, 75, dan 100. Model optimum yang diperoleh akan dievaluasi dengan confusion matriks yang terdiri atas akurasi, sensitivity, dan spesifisity.



Gambar 1 Flowchart Penelitian

2.3 Normalisasi Min-Max

Normalisasi digunakan untuk mengubah data yang memiliki rentang nilai yang cukup besar sehingga dapat menghasilkan model yang lebih akurat. Dalam normalisasi, data akan diubah namun informasi yang terkandung dalam data tidak berubah [11]. Salah satu metode normalisasi yang umum digunakan adalah MinMax Normalization, yang mengubah rentang nilai data menjadi antara 0 dan 1. Metode ini dapat digunakan dalam penelitian untuk mengoptimalkan kinerja model yang digunakan. Persamaan untuk menghitung MinMax Normalization dapat dilihat pada Persamaan 2 [12].

$$x' = \frac{x_i - \min(x)}{\max(x) - \min(x)} \tag{2}$$

Keterangan:

x' = Hasil normalisasi.

x_i = Nilai data.

min = Min data.

max = Max data.

2.4 Smote

Synthetic Minority Over-sampling Technique (SMOTE) merupakan teknik Oversampling menyeimbangkan jumlah distribusi dataset pada kelas minoritas dengan cara mensintesis dataset minoritas hingga jumlah dataset menjadi seimbang dengan jumlah dataset pada kelas mayoritas [13]. SMOTE akan membangkitkan data dari kelas minor dengan pendekatan ketetanggaan [14]. Tujuan dari penyeimbangan dataset adalah agar supaya setiap kelas mempunyai mempunyai frekuensi/jumlah data yang sama [15]. Pada penelitian ini untuk menyeimbangkan data menggunakan teknik yang disebut dengan Synthetic Minority dimana misal diberikan data dengan jumlah variabel p maka jarak antara

$$x^T = [x_1, x_2, \dots, x_p] \text{ dan } z^T = [z_1, z_2, \dots, z_p] \text{ adalah } d(x, y) = \sqrt{(x_1 - z_1)^2 + (x_2 - z_2)^2 + \dots + (x_p - z_p)^2}$$

atau biasa disebut sebagai jarak euclidean [23]. Ketika data bertipe numerik dan kategorik, perhitungan jarak tetap menggunakan jarak Euclidean namun menggunakan nilai median dari simpangan baku peubah numerik sebagai selisih nilai peubah kategorik. Nilai median ini dihitung ketika nilai kategori amatan dan tetangga terdekatnya berbeda [16]. Rumus perhitungan jarak Euclidean yaitu:

- a. Data Numerik Data yang memiliki jarak terdekat kemudian digunakan untuk membangkitkan data sintesis baru berdasarkan persamaan berikut :

$$X_{baru} = x + (x^* - x) \times Rand [0,1] \quad (3)$$

dengan

X_{baru} = data sintesis hasil dari replikasi

x = data yang akan direplikasi

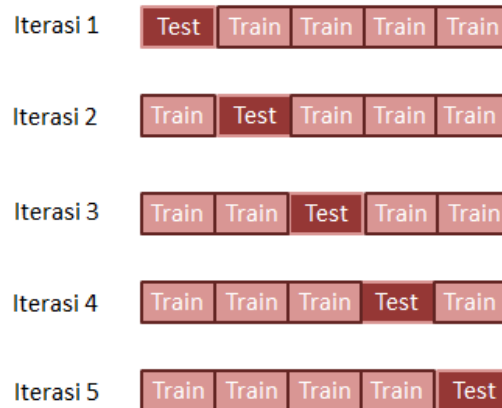
x^* = data yang memiliki jarak terdekat dari data yang akan di replikasi

$Rand[0,1]$ = bilangan acak antara 0 sampai 1

- b. Data kategorik Pada data kategorik, data buatan dibangkitkan dengan menentukan kategori yang paling sering muncul diantara amatan dengan k-tetangga terdekat (modus). Apabila terdapat kesamaan nilai dari kategorinya, maka akan dipilih secara acak [16].

2.5 K-Fold Cross Validation

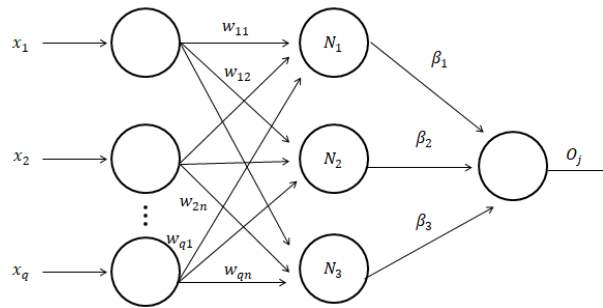
Crossvalidation atau dapat disebut estimasi rotasi adalah sebuah teknik validasi model untuk menilai bagaimana hasil statistik analisis akan menggeneralisasi kumpulan data independen. Salah satu teknik dari validasi silang adalah *k-fold cross validation* [17]. *K-fold cross validation* merupakan salah satu dari teknik yang difungsikan untuk memilah data menjadi data train serta data test [18]. Contoh simulasi dalam *Cross Validation* seperti pada Gambar 2.



Gambar 2 K-Fold Cross Validation

2.6 Extreme Learning Machine

Extreme Learning Machine (ELM) adalah salah satu jenis jaringan saraf tiruan yang hanya memiliki satu lapisan tersembunyi. ELM muncul sebagai pendekatan baru dalam pembelajaran jaringan saraf tiruan, berdasarkan perkembangan jaringan saraf tiruan feedforward dan menggunakan Single Layer Feedforward Neural Network (SLFN) [9]. ELM digunakan untuk mengatasi kelemahan dalam jaringan saraf tiruan feedforward, terutama dalam kondisi kecepatan pembelajaran (learning speed) yang lambat. ELM dapat bekerja ribuan kali lebih cepat daripada metode Backpropagation dan menghasilkan kinerja generalisasi yang lebih baik [19].



Gambar 3 Arsitektur ELM

Langkah-langkah algoritma ELM adalah sebagai berikut [20]:

- a. Inisialisasi bobot dan bias awal. Nilai bobot (W_{jk}) dihasilkan secara acak dalam rentang -1 hingga 1, dengan j adalah jumlah neuron tersembunyi dan q adalah jumlah neuron input, sedangkan nilai bias (b) dihasilkan secara acak dalam rentang 0 ke 1 untuk $1 \times j$.
- b. Hitung matriks keluaran lapisan tersembunyi menggunakan Persamaan 4 :

$$H_{init} = X_{(i)}W_{(i)}T + b_{(i)} \tag{4}$$

Keterangan:

$X(i)$ = data input ke- i

$W(i)T$ = bobot input yang di-transpose ke- i

$b(i)$ = nilai bias ke- i

Selanjutnya menghitung matriks H_{init} menggunakan fungsi aktivasi sigmoid biner. Hasil dari proses ini adalah pemetaan matriks dari matriks H_{init} yang diperoleh dari fungsi aktivasi sigmoid biner, yang merupakan fungsi nonlinier dengan nilai antara 0 dan 1. Perhitungan proses aktivasi sigmoid biner diperoleh pada Persamaan 5:

$$H = \frac{1}{1 + \exp(-H_{init})} \tag{5}$$

Di mana:

$X(i)$ = data masukan untuk sampel ke- i

$W(i)T$ = transpos bobot input untuk sampel ke- i

$b(i)$ = nilai bias untuk sampel ke- i

- c. Hitung invers Moore-Penrose menggunakan Persamaan 6:

$$H^\dagger = (H^T \cdot H)^{-1} \cdot H^T \tag{6}$$

H^\dagger = Matriks Moore-Penrose

H = Matriks output hiddenlayer

HT = Matriks output hidden layer yang ditranspose

- d. Menghitung bobot akhir (Persamaan 7).

$$\hat{B} = H^\dagger T \tag{7}$$

Keterangan:

$\hat{\beta}$ = Matriks output bobot

H^\dagger = Matriks Moore-Penrose

T = Matriks kelas data latih

- e. Menghitung hasil klasifikasi (Persamaan 8) .

$$\hat{Y} = H\hat{B} \tag{8}$$

Keterangan:

\hat{Y} = Hasil klasifikasi

H = Matriks output hidden layer

$\hat{\beta}$ = Matriks output bobot Proses ELM ini dilakukan

2.7 Confusion Matrix

Pada penelitian ini, akurasi hasil diuji dengan menggunakan Confusion Matrix. Confusion matrix adalah metode yang digunakan untuk menghitung akurasi konsep data mining. Akurasi atau keyakinan adalah laporan kasus prediksi yang positif, serta benar-benar positif untuk data aktual [21].

Tabel 2 Confusion Matrix

		Aktual	
		True	False
Prediksi	True	True Positive (TP)	False Positive (FP)
	False	False Negative (FN)	True Negative (TN)

Tabel matriks kebingungan terdiri dari empat nilai [22]:

- TP, True Positive, yaitu jumlah data positif yang diklasifikasikan dengan benar oleh sistem.
- TN, True Negative, yaitu jumlah data negatif yang diklasifikasikan dengan benar oleh sistem.
- FN, False Negative, yaitu jumlah data negatif tetapi salah klasifikasi oleh sistem.
- FP, False Positive, yaitu jumlah data positif tetapi diklasifikasikan secara salah oleh sistem.

Pengujian merupakan langkah terakhir dalam mengevaluasi sistem [23]. Akurasi adalah rasio prediksi yang benar (positif dan negatif) terhadap total data. Dengan kata lain, ini membandingkan nilai prediksi dengan nilai sebenarnya. Nilai akurasi dapat diperoleh dengan menggunakan persamaan 9:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (9)$$

Nilai presisi menggambarkan banyaknya data kategori positif yang terklasifikasi benar dibagi dengan total data yang terklasifikasi positif. Presisi dapat diperoleh dengan menggunakan persamaan 10:

$$Presisi = \frac{TP}{TP + FP} \times 100\% \quad (10)$$

Sementara itu, recall menunjukkan persentase data kategori positif yang diklasifikasikan dengan benar oleh sistem:

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (11)$$

Error adalah kasus yang teridentifikasi secara tidak benar pada sejumlah data, sehingga dapat dilihat seberapa tinggi tingkat error pada sistem yang digunakan. Persentase error dapat dihitung dengan menggunakan persamaan berikut:

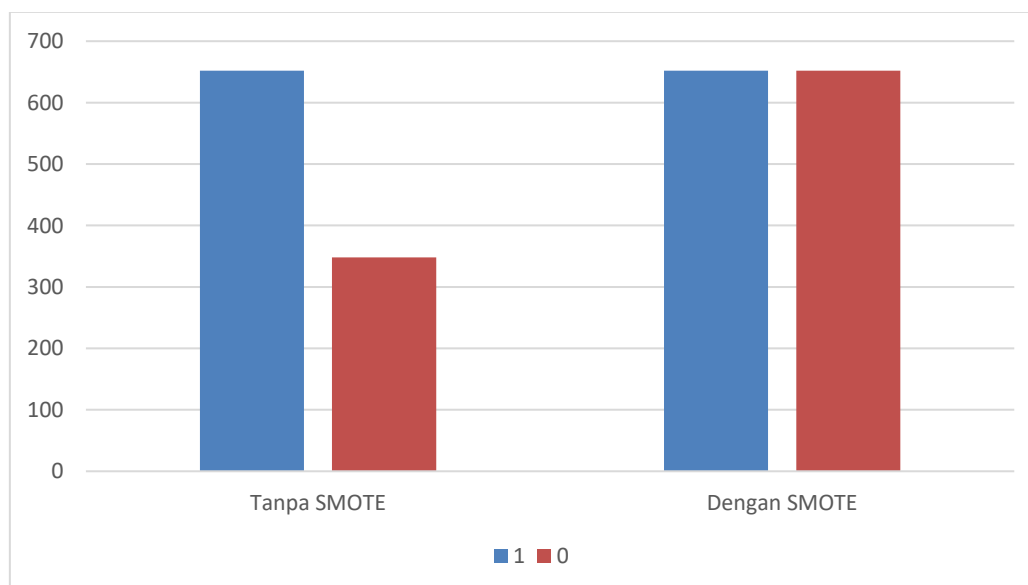
$$Error = \frac{FP}{TP} \times 100\% \quad (12)$$

3. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan data dari web kaggle. Dari data tersebut terdapat data yang menyimpang dari data yang lainnya atau disebut dengan data outlier sehingga harus diatasi dengan menggunakan persamaan (1). Setelah itu, data dinormalisasikan menggunakan persamaan (2) agar tetap konstan. Data hasil normalisasi ditunjukkan oleh Tabel 4. Dari data tersebut jumlah data antara kelas 1 dan 0 tidak seimbang atau dapat disebut dengan imbalance class dengan jumlah kelas 1 sebanyak 652 dan kelas 0 sebanyak 348. Oleh karena itu perlu diatasi dengan menggunakan metode SMOTE. Hasil dari resampling menggunakan SMOTE ditunjukkan oleh Gambar 2.

Tabel 3 Hasil Preprocessing

NO	Aluminium	Amonia	...	Potabilitas
1.	0.3300	0.3038	...	1
2.	0.4640	0.7089	...	1
⋮	⋮	⋮	⋮	⋮
8000.	0.0040	0.1449	...	1



Gambar 4 Hasil Smote

Metode yang digunakan dalam klasifikasi ini adalah Extreme Learning Machine (ELM). Langkah awal yaitu pembagian data training dan data testing menggunakan metode k-fold cross validation dengan partisi k yang digunakan sebesar 5 dan 10. Hasil dari nilai k kemudian dibandingkan untuk dicari nilai k yang paling optimal. Pada penelitian ini, melakukan beberapa uji coba terhadap hidden node dan penggunaan atau tanpa SMOTE. Nilai uji coba hidden node yang digunakan adalah 5, 25, 50, 75, dan 100. Hasil uji coba ditunjukkan oleh Tabel 4 dan 5.

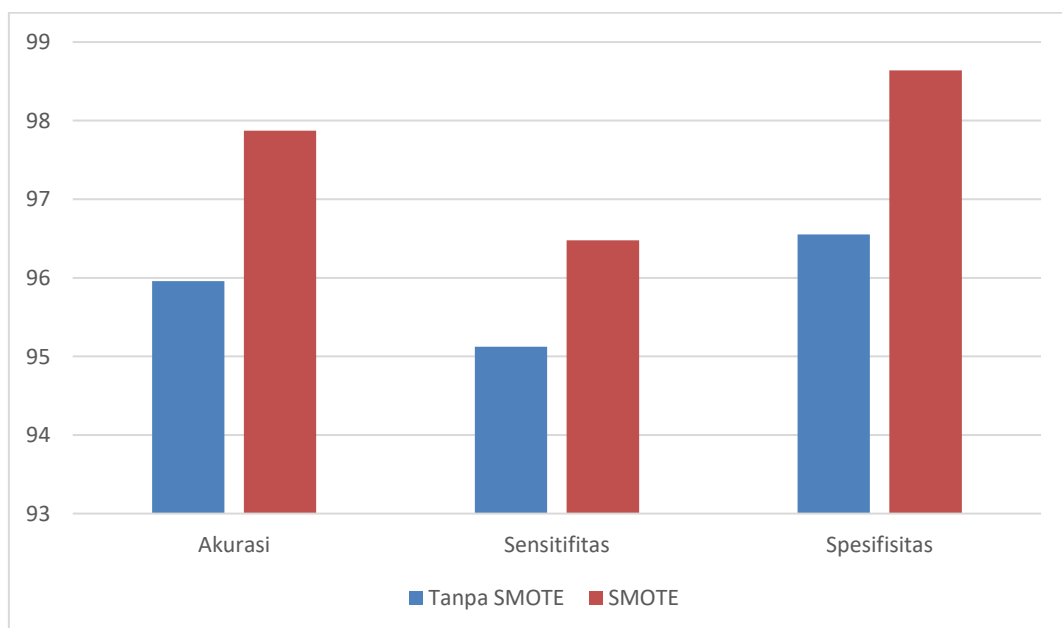
Tabel 4 Hasil Uji Coba Dengan Smote

k-fold	node hidden	Akurasi	Sensitivitas	Spesifitas
5	5	67.3750	29.1139	87.5836
	25	89.7934	84.0868	92.8161
	50	97.0496	96.5642	97.3077
	75	96.3591	96.2025	96.4423
	100	96.7357	94.7464	97.7906
10	5	61.6250	33.1690	93.7984
	25	93.0818	93.3099	92.9550
	50	96.1055	95.4225	96.4844
	75	97.1106	96.4789	97.4609
	100	97.8723	96.4789	98.6408

Tabel 5 Hasil Uji Coba Tanpa Smote

k-fold	node hidden	Akurasi	Sensitivitas	Spesifitas
5	5	67.0000	23.9437	90.6977
	25	89.9497	88.7324	90.6250
	50	92.9648	92.9577	92.9688
	75	92.0000	87.3239	94.5736
	100	92.4623	90.1408	93.7500
10	5	61.0000	24.3902	86.4407
	25	91.0000	85.3659	94.9153
	50	91.9192	90.2439	93.1034
	75	92.9293	90.2439	94.8276
	100	95.9596	95.1220	96.5517

Berdasarkan Tabel 4 dan Tabel 5 diperoleh bahwa penggunaan SMOTE untuk mengatasi imbalance data sangat berpengaruh pada hasil uji coba. Diperoleh hasil terbaik penggunaan metode SMOTE yaitu pada k-fold 10 dan node hidden 100 menghasilkan akurasi 97.873%, sensitifitas 96.4789%, dan spesifisitas 98.6408%. Hal tersebut membuktikan bahwa metode SMOTE berpengaruh terhadap hasil uji coba. Grafik perbandingan hasil uji coba terbaik pada penggunaan metode SMOTE dan tanpa SMOTE dengan k-fold 10 dan node hidden 100 pada ditunjukkan oleh Gambar 4.



Gambar 5 Evaluasi Kinerja Dengan Smote dan Tanpa Smote

4. KESIMPULAN

Berdasarkan hasil penelitian di atas diperoleh kesimpulan bahwa penerapan Smote pada metode ELM untuk mengatasi imbalance data memberikan hasil yang baik saat menangani ketidakseimbangan kelas pada dataset kualitas air. Nilai evaluasi kerja untuk data dengan smote memiliki hasil terbaik saat menggunakan $K=10$ dengan nilai akurasi 97.8723% sedangkan untuk data tanpa smote memiliki hasil terbaik saat $K=10$ dengan nilai akurasi 95.9596% yang artinya klasifikasi pada data imbalance lebih efektif saat menggunakan Smote untuk meningkatkan performa klasifikasi. Diharapkan hasil klasifikasi ini dapat membantu perencanaan strategi dalam menjaga kualitas, kuantitas, dan kontinuitas air bersih.

5. SARAN

Penelitian selanjutnya dapat memperluas dengan memasukkan proses normalisasi data selain min-max guna untuk melihat apakah akan menghasilkan evaluasi kinerja yang lebih baik atau tidak. Serta dapat menggunakan metode yang lebih kompleks untuk memperoleh hasil akurasi yang lebih baik lagi.

DAFTAR PUSTAKA

- [1] P. Pattasang and Rika Ampuh Hadiguna, "Rancangan Usaha Air Minum Dalam Kemasan Menggunakan Merek Mineral Santry (Di Pondok Pesantren Insan Mandiri Batam)," *J. Manaj. Pendidik. Dan Ilmu Sos.*, vol. 2, no. 2, pp. 705–712, 2021, doi: 10.38035/jmpis.v2i2.630.
- [2] J. T. Sipil *et al.*, "Analisa Sistem Distribusi Air Bersih Pada Perumahan Mulawarman Residence Di," *J. Ilm. Univ. Batanghari Jambi*, vol. 20, no. 3, p. 1075, 1945.
- [3] M. S. Dr. Rahmatsyah, M. S. Dr. Rita Juliani, M. S. Drs. Nusyirwan, and M. P. Rajo Hasim Lubis, *Fisika Lingkungan*. Bandung: Penerbit Media Sains Indonesia, 2021.
- [4] P. Studi, K. Masyarakat, F. I. Kesehatan, and U. P. N. V. Jakarta, "ANALISIS HUBUNGAN KONDISI EKONOMI MASYARAKAT TERHADAP PEMAKAIAN AIR DOMESTIK Keren Stelin Maliangkay 1 , Nabila Erinaputri 2 , Nadya Salsabila 3 , Marina Ery Setiawati 4," pp. 592–599, 2022.
- [5] H. Said, N. Matondang, H. Nurramdhani Irmada, and S. Informasi, "Penerapan Algoritma K-Nearest Neighbor Untuk Memprediksi Kualitas Air Yang Dapat Dikonsumsi Application of K-Nearest Neighbor Algorithm to Predict Consumable Water Quality," vol. 21, no. 2, pp. 256–267, 2022, [Online]. Available: www.kaggle.com.
- [6] I. A. Sukamari, B. Kwaji, J. Alheri, and I. E. Abia, "Application of Water Quality Index to Assess the Potability of Some Domestic Water Supply Sources in Mubi North , Nigeria," *Int. J. Eng. Res. Technol.*, vol. 9, no. 08, pp. 848–860, 2020.
- [7] S. Mohd-Asharuddin, N. Zayadi, W. Rasit, and N. Othman, "Water Quality Characteristics of Sembrong Dam Reservoir, Johor, Malaysia," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 136, no. 1, 2016, doi: 10.1088/1757-899X/136/1/012058.
- [8] Y. Romdania, A. Herison, G. E. Susilo, and E. Novilyansa, "KAJIAN PENGGUNAAN METODE IP, STORET, dan CCME WQI DALAM MENENTUKAN STATUS KUALITAS AIR," *J. Spat. Wahana Komun. dan Inf. Geogr.*, vol. 18, no. 2, pp. 133–144, 2018, doi: 10.21009/spatial.182.07.
- [9] R. J. Rumandan, R. Nuraini, N. Sadikin, and Y. Rahmanto, "Klasifikasi Citra Jenis Daun Berkhasiat Obat Menggunakan Algoritma Jaringan Syaraf Tiruan Extreme Learning Machine," *J. Comput. Syst. Informatics*, vol. 4, no. 1, pp. 145–154, 2022, doi: 10.47065/josyc.v4i1.2586.
- [10] H. Pratiwi and K. Harianto, "Perbandingan Algoritma ELM Dan Backpropagation

- Terhadap Prestasi Akademik Mahasiswa,” *J-SAKTI (Jurnal Sains Komput. dan Inform.*, vol. 3, no. 2, p. 282, 2019, doi: 10.30645/j-sakti.v3i2.147.
- [11] Sholeh, D. Andayati, and Y. Rachmawati, “Data Mining Model Klasifikasi Menggunakan K-Nearest Neighbor With Normalization For Diabetes Prediction,” *Telka*, vol. 12, no. 1, pp. 77–87, 2022.
- [12] I. Permana and F. Nur Salisah, “The Effect of Data Normalization on the Performance of the Classification Results of the Backpropagation Algorithm,” *Ijrse*, vol. 2, no. 1, pp. 67–72, 2022, [Online]. Available: <https://media.neliti.com/media/publications/485639-pengaruh-normalisasi-data-terhadap-perfo-e19e3a00.pdf>.
- [13] G. Gumelar *et al.*, “Kombinasi Algoritma Sampling dengan Algoritma Klasifikasi untuk Meningkatkan Performa Klasifikasi Dataset Imbalance,” *Pros. Semin. Nas. Sist. Inf. dan Teknol. ke 5*, pp. 250–255, 2021.
- [14] M. Syukron, R. Santoso, and T. Widiharih, “Perbandingan Metode Smote Random Forest Dan Smote Xgboost Untuk Klasifikasi Tingkat Penyakit Hepatitis C Pada Imbalance Class Data,” *J. Gaussian*, vol. 9, no. 3, pp. 227–236, 2020, doi: 10.14710/j.gauss.v9i3.28915.
- [15] M. Ibnu and C. Rachmatullah, “Penerapan SMOTE untuk Meningkatkan Kinerja Klasifikasi Penilaian Kredit,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 10, no. 1, pp. 2407–389, 2023, doi: 10.30865/jurikom.v10i1.5612.
- [16] F. N. Umma, B. Warsito, and D. A. I. Maruddani, “Klasifikasi Status Kemiskinan Rumah Tangga Dengan Algoritma C5.0 Di Kabupaten Pematang,” *J. Gaussian*, vol. 10, no. 2, pp. 221–229, 2021, doi: 10.14710/j.gauss.v10i2.29934.
- [17] H. Azis, P. Purnawansyah, F. Fattah, and I. P. Putri, “Performa Klasifikasi K-NN dan Cross Validation Pada Data Pasien Pengidap Penyakit Jantung,” *Ilk. J. Ilm.*, vol. 12, no. 2, pp. 81–86, 2020, doi: 10.33096/ilkom.v12i2.507.81-86.
- [18] T. Ridwansyah, “Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier,” *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 2, no. 5, pp. 178–185, 2022, doi: 10.30865/klik.v2i5.362.
- [19] S. Multazam, I. Cholissodin, and S. Adinugroho, “Implementasi Metode Extreme Learning Machine pada Klasifikasi Jenis Penyakit Hepatitis berdasarkan Faktor Gejala,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 3, pp. 789–797, 2020, [Online]. Available: <http://j-ptiik.ub.ac.id>.
- [20] R. Ramadhan and W. F. Mahmudy, “Optimasi Bobot Awal Extreme Learning Machine menggunakan Algoritme Genetika untuk Klasifikasi Penanganan Human Papilloma Virus,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 6, pp. 5616–5623, 2019.
- [21] D. J. Lubis and G. K. Gusti, “Penerapan Algoritma Naïve Bayes Untuk Penentuan Balita Penerima Makanan Tambahan (PMT) Berdasarkan Status Gizi Di Pos Pelayanan Terpadu (POSYANDU),” *TeknoIS J. Ilm. Teknol. Inf. dan Sains*, vol. 13, no. 1, pp. 58–66, 2023, doi: 10.36350/jbs.v13i1.177.
- [22] B. P. Pratiwi, A. S. Handayani, and S. Sarjana, “Pengukuran Kinerja Sistem Kualitas Udara Dengan Teknologi Wsn Menggunakan Confusion Matrix,” *J. Inform. Upgris*, vol. 6, no. 2, pp. 66–75, 2021, doi: 10.26877/jiu.v6i2.6552.
- [23] A. Rini and H. Aprianto, “Geographic Information System of Health Service Place in Palembang,” in *Journal of Physics: Conference Series*, 2019, vol. 1167, no. 1: IOP Publishing, p. 012065.