

Performa Metode Klasifikasi Tunggal dan *Ensemble Model* dalam Identifikasi Baku Mutu Air

Prasetya Widiharso^{1*}, Siti Sendari², Anik Nur Handayani³, Nastiti Susetyo Fanani Putri⁴

^{1, 2, 3, 4}Program Studi Teknik Elektro, Universitas Negeri Malang

^{1,2,3,4}Jl. Semarang No.5 Kota Malang, 65145, Indonesia

E-mail: prasetya.widiharso.2005348@students.um.ac.id¹, siti.sendari.ft@um.ac.id², aniknur.ft@um.ac.id³, nastiti.susetyo.2005348@students.um.ac.id⁴

Abstrak

Info Naskah:

Naskah masuk: 29 Mei 2022

Direvisi: 15 Juni 2022

Diterima: Juli 2022

Pengklasifikasian kualitas air untuk kebutuhan sarana rekreasi, peternakan, perikanan, dan pertanian diperlukan guna menentukan pemanfaatan berdasarkan kualitas air sesuai baku mutu air nasional. Metode yang digunakan dalam klasifikasi pemanfaatan kualitas air adalah metode *K-Nearest Neighbors (KNN)*, *Support Vector Machine (SVM)*, *Naïve Bayes (NB)*, dan *Ensemble Model*. Parameter yang diukur terdiri dari temperatur, TDS, TSS, pH, COD, BOD, DO, dan curah hujan. Tujuan utama dari penelitian adalah mengetahui performa metode klasifikasi tunggal dan *ensemble model* terhadap jenis data dengan distribusi kelas yang tidak seimbang. Obyek klasifikasi peruntukan air sungai dan waduk dibagi kedalam dua kelas. Pertama, kelas untuk peruntukan sarana rekreasi, perikanan, dan peternakan. Kedua, kelas untuk peruntukan budidaya pertanian. Hasil uji akurasi penerapan metode *KNN* diperoleh hasil 86%, *SVM* diperoleh hasil 87%, *NB* diperoleh hasil 90,57%. Sedangkan melalui *Ensemble Model* diperoleh hasil *Bagging Classifier* 94,43%, *Gradient Boosting Classifier* 94,96%, *Adaboost Classifier* 95,94%.

Abstract

Keywords:

air;

k-nearest neighbors;

support vector machine;

naïve bayes;

ensemble model.

Water quality classification for the needs of recreational facilities, livestock, fisheries, and plantations is needed to determine utilization based on water quality according to national water quality standards. The methods used in this research are *K-Nearest Neighbors (KNN)*, *Support Vector Machine (SVM)*, *Naïve Bayes (NB)*, and *Ensemble Model*. The parameters measured consisted of temperature, TDS, TSS, pH, COD, BOD, DO, and rainfall. The main objective of this research is to discover the performance of a single classification method and ensemble model on data types with unbalanced class distributions. Classification objects are divided into two classes. First, is the class for the designation of recreational facilities, fisheries, and livestock. Second, the class for the allotment of crop cultivation. The test results of the application of the *KNN* obtained 86%, *SVM* obtained 87%, and *NB* obtained 90.57%. Meanwhile, through the ensemble model, the results obtained are 94.43% *Bagging Classifier*, 94.96% *Gradient Boosting Classifier*, and 95.94% *Adaboost Classifier*.

*Penulis korespondensi:

Prasetya Widiharso

E-mail: prasetya.widiharso.2005348@students.um.ac.id

1. Pendahuluan

Air sungai dan waduk dimanfaatkan sebagai air minum, sarana rekreasi air, perikanan, peternakan dan budidaya pertanian, namun air dari sumber tersebut seringkali terkontaminasi logam berat bersumber dari air limbah yang kurang tepat pengelolaannya. Pemanfaatan air sungai dan waduk memerlukan pengukuran dan pengklasifikasian kedalam kelas peruntukannya sehingga tidak menimbulkan gangguan kesehatan bagi manusia maupun makhluk hidup lain. Konsentrasi logam berat yang melebihi normal dalam tubuh dapat mempengaruhi kesehatan manusia. Berdasarkan alasan tersebut, maka pengamatan kualitas air sungai dan waduk sangat penting dilakukan[1].

Pengukuran tingkat pencemaran di sungai dan waduk yang telah ada dilakukan dengan mengukur kandungan oksigen terlarut dengan titrasi dan pengukuran dengan DO meter elektronik. Pengukuran tersebut mudah dilakukan namun harus bersentuhan langsung dengan objek yang akan diukur. Metode titrasi dilakukan melalui analisis kimia dengan mengukur konsentrasi dan reaktan. Proses titrasi meliputi titran sebagai larutan yang diketahui dan titrat sebagai larutan yang titrasinya telah ditentukan[2].

Beberapa penelitian terkait lain yang telah ada diantaranya adalah Aryasa, dkk. (2016) [3], penelitian ini bertujuan untuk membandingkan metode *Particle Swarm Optimization (PSO)*, *Support Vector Machine (SVM)*, dan hybrid SVM-PSO dalam otomatisasi buka tutup pintu pembuangan limbah nikel. Hasil yang diperoleh metode *hybrid SVM-PSO* lebih unggul dibandingkan *PSO* dan *SVM* dalam analisis kualitas baku mutu air pembuangan limbah nikel.

Haghiabi, dkk (2018) [4], tujuan dari penelitian ini untuk mempelajari metode Jaringan Syaraf Tiruan (*ANN*), *Group Method Of Data Handling (GMDH)*, dan *Support Vector Machine* dalam memprediksi komponen kualitas air Sungai Tireh Iran. Pengujian melalui beberapa jenis kernel berbeda digunakan dalam penelitian. Hasil *ANN* dan *SVM* menunjukkan bahwa kedua model memiliki kinerja yang sesuai untuk memprediksi komponen kualitas air. Dalam penelitian, tansig dan *RBF* diketahui menjadi fungsi transfer dan kernel yang berkinerja terbaik. Noviani, dkk (2020) [5], tujuan dari penelitian ini adalah untuk pengklasteran keluhan dari pelanggan air dengan solusi yang akan diberikan oleh divisi technical service PT.Suryacipta Swadaya. Metode yang digunakan adalah *K-Medoids*. Hasil yang didapatkan, divisi technical service PT.Suryacipta Swadaya memperoleh peningkatan efektifitas kinerja melalui pengelompokan permasalahan dan solusi yang harus dijalankan, dibandingkan dengan sebelum menerapkan metode *K-Medoids*.

Yulias, dkk (2021) [6], penelitian ini menerapkan metode algoritma *Naive Bayes* yang digunakan untuk menentukan pilihan sumber air baku guna pembangunan fasilitas air minum baru. Hasil penerapan metode *Naive Bayes* melalui *confusion matrix* menunjukkan tingkat akurasi yang baik.

Berdasarkan penelitian yang telah ada, beberapa peneliti tersebut menggunakan metode klasifikasi tunggal yang dioptimasi untuk mendapatkan peningkatan performa dalam prediksi. Dalam penelitian ini, mengusulkan perbandingan metode klasifikasi tunggal yang terdiri dari *K-*

Nearest Neighbors, *Support Vector Machine*, *Naive Bayes* dengan *ensemble model* yang terdiri dari *Bagging Classifier*, *Gradient Boosting Classifier*, dan *Adaboost Classifier*. Metode klasifikasi tunggal dan *ensemble model* tersebut diuji performanya dalam menangani distribusi data yang tidak seimbang berdasarkan kondisi data riil yang diperoleh dari sumber. Parameter yang diukur terdiri dari temperatur, total padatan terlarut (TDS), padatan tersuspensi (TSS), tingkat keasaman (pH), oksigen biokimiawi (BOD), oksigen kimiawi (COD), oksigen terlarut (DO), dan curah hujan. Kelas atau label sebagai target dibagi menjadi dua, dua kelas tersebut terdiri dari kelas untuk peruntukan sarana rekreasi, perikanan, dan peternakan, diberikan label 1. Kedua, kelas untuk peruntukan budidaya pertanian, diberikan label 0. Pembagian kelas berdasar nilai tiap parameter *input* bersumber dari data sekunder yaitu ketentuan nilai baku mutu air nasional.

2. Metode Penelitian

2.1 Algoritma Machine Learning

Algoritma *machine learning* bertujuan untuk membuat aturan dalam melatih komputer guna memahami model berdasarkan data yang dimiliki. Jenis algoritma *machine learning* diantaranya adalah, *supervised learning*, *unsupervised learning*, *semi supervised learning*, dan *reinforcement learning*[7]. *Supervised learning* adalah metode pembelajaran dengan menggunakan data latih dalam proses pembelajaran. Data latih tersebut merupakan data yang telah diberikan label. Beberapa algoritma *supervised learning* adalah Regresi Linier Berganda, *Decision Tree*, *Random Forest*, *Naive Bayes*, *Nearest Neighbors*, *Support Vector Machine*, dan *Artificial Neural Network* [7][8][9][10][11].

Pada teknik *unsupervised learning* tidak diperlukan data latih dalam pembelajaran, sistem mempelajari data tak berlabel dan mengelompokkan berlandaskan fitur yang menyertai. Beberapa algoritma *unsupervised learning* adalah *K-Means*, *DbSCAN*, *Fuzzy C-Means*, dan *Self Organizing Map*. Teknik *semi supervised*, merupakan teknik kombinasi antara *supervised* dan *unsupervised learning*. *Reinforcement Learning* adalah metode penerapan melalui agen cerdas untuk menyesuaikan dengan kondisi lingkungan dari sistem[7][8][9].

2.2 Support Vector Machine (SVM)

SVM merupakan metode untuk menyelesaikan masalah klasifikasi baik linier maupun non-linier. *SVM* bekerja dengan mencari *hyperplane* yang optimal dengan memaksimalkan jarak antar kelas yang dimiliki data. Dalam fungsi 2 dimensi yang digunakan untuk mengklasifikasikan antar kelas disebut garis, dalam fungsi 3 dimensi disebut sebagai bidang, dan dalam kelas berdimensi lebih tinggi disebut *hyperplanes* [8][9][10][12].

2.3 K-Nearest Neighbors (KNN)

Algoritma *K-nearest Neighbor (KNN)* bekerja dengan menghitung jarak data baru dari suatu objek menggunakan data referensi atau data latih terdekat. Nilai *K* merepresentasikan jumlah tetangga terdekat yang ditentukan untuk menentukan prediksi label kelas dari data uji. Selanjutnya berdasarkan nilai *K* yang telah dipilih, diambil *voting* berdasarkan frekuensi keanggotaan kelas terbanyak

dalam himpunan K tetangga. *KNN* adalah algoritma pembelajaran terawasi yang mengklasifikasikan hasil instance kueri baru berdasarkan sebagian besar kelas dalam kategori dalam data [8][9][10][12][13].

2.4 Naive Bayes (NB)

Naive Bayes adalah metode turunan dari teorema Bayes. Teorema Bayes adalah metode menghitung probabilitas bersyarat yang diperkenalkan oleh Thomas Bayes. Metode *NB* bekerja untuk memprediksi peluang masa depan berdasarkan peristiwa masa lalu[7][8][9][10][11].

2.5 Ensemble Model

Terdapat 4 cara untuk meningkatkan performa *machine learning* diantaranya adalah, peningkatan performa dengan data, algoritma, *tuning*, dan *ensembles*[7]. *Ensemble learning* merupakan suatu teknik algoritma untuk mempelajari data melalui kombinasi algoritma guna mendapatkan hasil akurasi yang lebih baik dibandingkan dengan penggunaan algoritma model tunggal. Beberapa metode *ensemble* adalah *bagging*, *boosting*, dan *stacking*[8][9][14].

Bagging atau *bootstrap aggregating*, teknik memanfaatkan beberapa model algoritma yang sama, selanjutnya melatih setiap model pada sampel berbeda dari *dataset* yang sama. Prediksi yang diperoleh dari setiap model digabungkan dan diambil *voting* atau rata-rata [8][9][14]. *Boosting* adalah varian *bagging*, di mana model tunggal yang dibuat ditempatkan berulang kali secara berurutan. Semua titik data yang salah diklasifikasikan oleh model sebelumnya dikoreksi dalam pemrosesan model berikutnya. Hal ini dilakukan untuk meningkatkan akurasi model secara keseluruhan dengan merata-ratakan semua hasil prediksi [8][9][14].

Stacking adalah suatu metode yang menggunakan metode *klasifier* dasar pada proses *training*. Tahapan awal dalam membangun model yaitu metode *klasifier* dasar dilatih dengan *dataset* yang sama untuk menghasilkan masing-masing prediksi. Pada tahapan selanjutnya *meta klasifier* mengambil hasil prediksi dari *klasifier* dasar sebagai *input* untuk menentukan kelas yang tepat terhadap data *training*[15][16].

2.6 Data dan Labeling

Tahapan dari penelitian seperti yang diperlihatkan pada Gambar 1, diawali dengan pengumpulan data primer dan sekunder, penentuan fitur dan label dari data, *preprocessing*, implementasi metode yang digunakan, evaluasi dan validasi. Obyek dari penelitian adalah data pengamatan air sungai dan waduk. Parameter yang diamati sebagai fitur input terdiri dari Temperatur, TDS, TSS, pH, COD, BOD, DO, dan curah hujan. Data Primer diambil dari Laboratorium Lingkungan I Perusahaan Umum Jasa Tirta. Lokasi titik pengamatan yang diamati terdiri dari Waduk Sutami Hilir, Cangkir Tambangan, Muara Kali Tengah. Sedangkan titik pengamatan klimatologi (curah hujan) diambil pada ARR Sutami, ARR Karang Pilang, dan ARR

Kedurus. Periode pengamatan yang digunakan bulan Januari 2019 sampai dengan bulan Desember 2021. Data Sekunder berupa ketentuan pengelompokan kelas baku mutu air diambil dari Peraturan Pemerintah Republik Indonesia Nomor 22 Tahun 2021 Tentang Penyelenggaraan Perlindungan dan Pengelolaan Lingkungan Hidup, diperlihatkan pada Tabel 2.

Berdasarkan data sekunder yang telah diperoleh diketahui kelas baku mutu air terdiri dari, kelas 1 dengan peruntukan baku mutu untuk air minum, kelas 2 dengan peruntukan sebagai prasarana atau sarana rekreasi air, pembudidayaan ikan air tawar, peternakan, dan mengairi pertanaman. Kelas 3 diperuntukkan sebagai pembudidayaan ikan air tawar, pengelolaan peternakan, serta mengairi pertanaman. Kelas 4 diperuntukkan hanya untuk mengairi pertanaman[17], seperti pada Tabel 1.

Tabel 1. Kelas baku mutu air dan peruntukannya[17]

Kelas	Peruntukkan
Kelas I	Air minum
Kelas II	Rekreasi air, budidaya ikan, peternakan, pertanaman
Kelas III	Budidaya ikan, peternakan, pertanaman
Kelas IV	Pertanaman

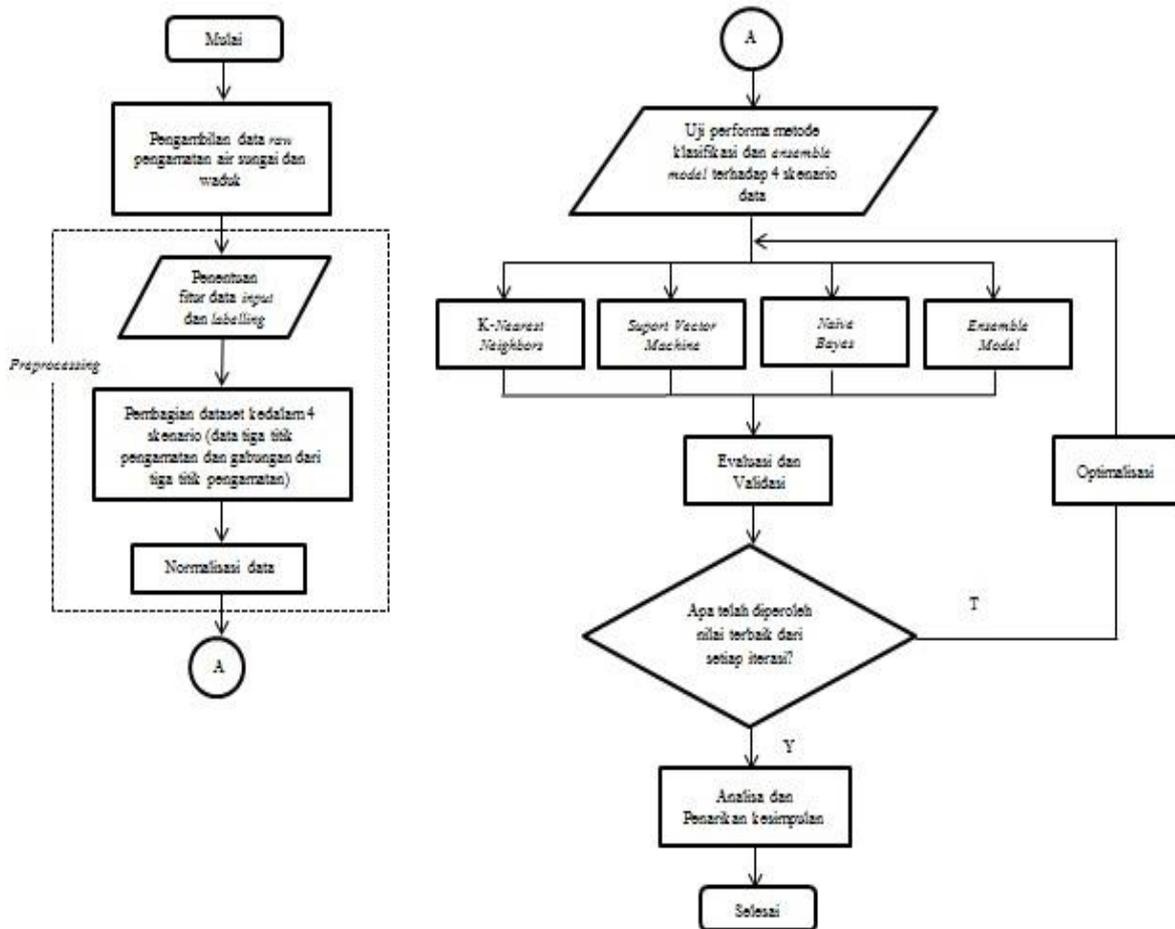
Tabel 2. Baku mutu air[18]

No	Parameter	Unit	Kelas				Limit
			I	II	III	IV	
1	Temperatur	°C	Dev3	Dev3	Dev3	Dev3	max
2	TDS	mg/L	1000	1000	1000	2000	max
3	TSS	mg/L	40	50	100	400	max
4	pH	-	6-9	6-9	6-9	6-9	max
5	BOD	mg/L	2	3	6	12	max
6	COD	mg/L	10	25	40	80	max
7	DO	mg/L	6	4	3	1	min

Dalam penelitian klasifikasi data air sungai dan waduk dibagi ke dalam dua kelas atau label. Dua kelas tersebut terdiri dari kelas untuk peruntukan sarana rekreasi, perikanan, dan peternakan, diberikan label 1. Kedua, kelas untuk peruntukan budidaya pertanaman, diberikan label 0.

2.7 Preprocessing

Preprocessing merupakan suatu cara mempermudah analisis data dengan mengubah data mentah ke dalam bentuk data yang lebih sederhana untuk dipahami. Tahapan dari *Preprocessing* diantaranya adalah *data cleaning*, data integrasi, transformasi data, dan *data reduction*[8]. Normalisasi data yang diterapkan adalah uji coba penggunaan *min-max scaler* dan *standard scaler*. Penerapan pembagian *dataset* menjadi data latih dan data tes sebesar 20%, 33%, serta penerapan *K-Fold Validation* pada *dataset*.



Gambar 1. Diagram alur penelitian

2.8 Implementasi metode

Pada bagian implementasi akan diterapkan beberapa metode klasifikasi tunggal diantaranya adalah *K-Nearest Neighbors*, *Support Vector Machine*, dan *Naive Bayes*. Pada langkah kedua yaitu penerapan *Ensemble Model* untuk mengetahui perbandingan peningkatan akurasi.

2.9 Evaluasi dan Validasi

Pada tahapan evaluasi dan validasi hasil dari masing-masing metode klasifikasi tunggal dilakukan pengujian ulang dengan *cross validation* pada data. Dalam penerapan *cross-validation*, setiap himpunan data yang telah dibagi rata digunakan beberapa kali untuk saling bertukar sebagai data uji dan data latih. Secara sederhananya metode *k-fold cross-validation* menerapkan pembagian data ke dalam k bagian dalam jumlah yang sama[19].

Evaluasi dan validasi merupakan suatu metode pengukuran kinerja model klasifikasi melalui perbandingan nilai aktual dan nilai prediksi. *Confusion Matrix* adalah matrik yang memetakan penghitungan performa pada *machine learning*, dimana keluaran dapat berupa dua kelas atau lebih. *Confusion Matrix* tersusun dari tabel matrik dengan susunan nilai prediksi dan nilai aktual. *Confusion Matrix* terdiri dari *True Positif*, *True Negatif*, *False Positif*, dan *False Negatif*[8].

	Temperatur	TDS	TSS	pH	BOD	COD	DO	CurahHujan	kelas
0	30.1	108.2	9.5	7.8	6.9	17.2	5.1	8.45	1
1	26.7	108.2	18.2	7.6	14.6	39.9	5.0	8.45	0
2	29.6	108.2	117.7	8.0	11.2	34.3	5.2	3.43	1
3	29.8	108.2	66.1	7.3	12.9	32.8	5.2	3.43	0
4	29.0	108.2	94.3	7.8	8.6	26.0	5.3	7.39	1

Gambar 2. Dataset pengamatan air sungai dan waduk

	Temperatur	TDS	TSS	pH	BOD	COD	DO	CurahHujan
0	0.954949	-1.154374	-0.656812	1.550118	-0.357483	-0.435253	0.914155	0.675926
1	-1.598869	-1.154374	-0.592058	0.968312	-0.168061	-0.304092	0.848847	0.675926
2	0.579387	-1.154374	0.148530	2.131925	-0.251702	-0.336449	0.979462	-0.168535
3	0.729612	-1.154374	-0.235534	0.095602	-0.209882	-0.345116	0.979462	-0.168535
4	0.128714	-1.154374	-0.025638	1.550118	-0.315663	-0.384407	1.044770	0.497613

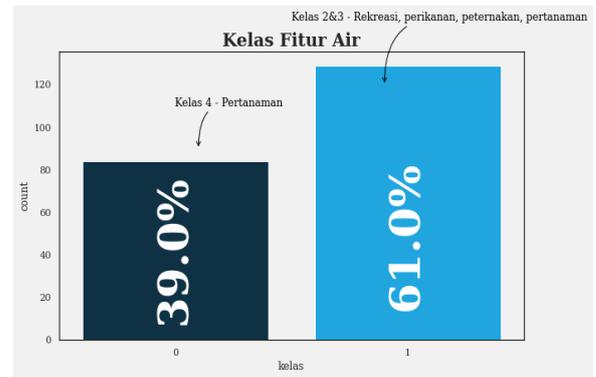
Gambar 3. Hasil normalisasi dengan Z-score

3. Hasil dan Pembahasan

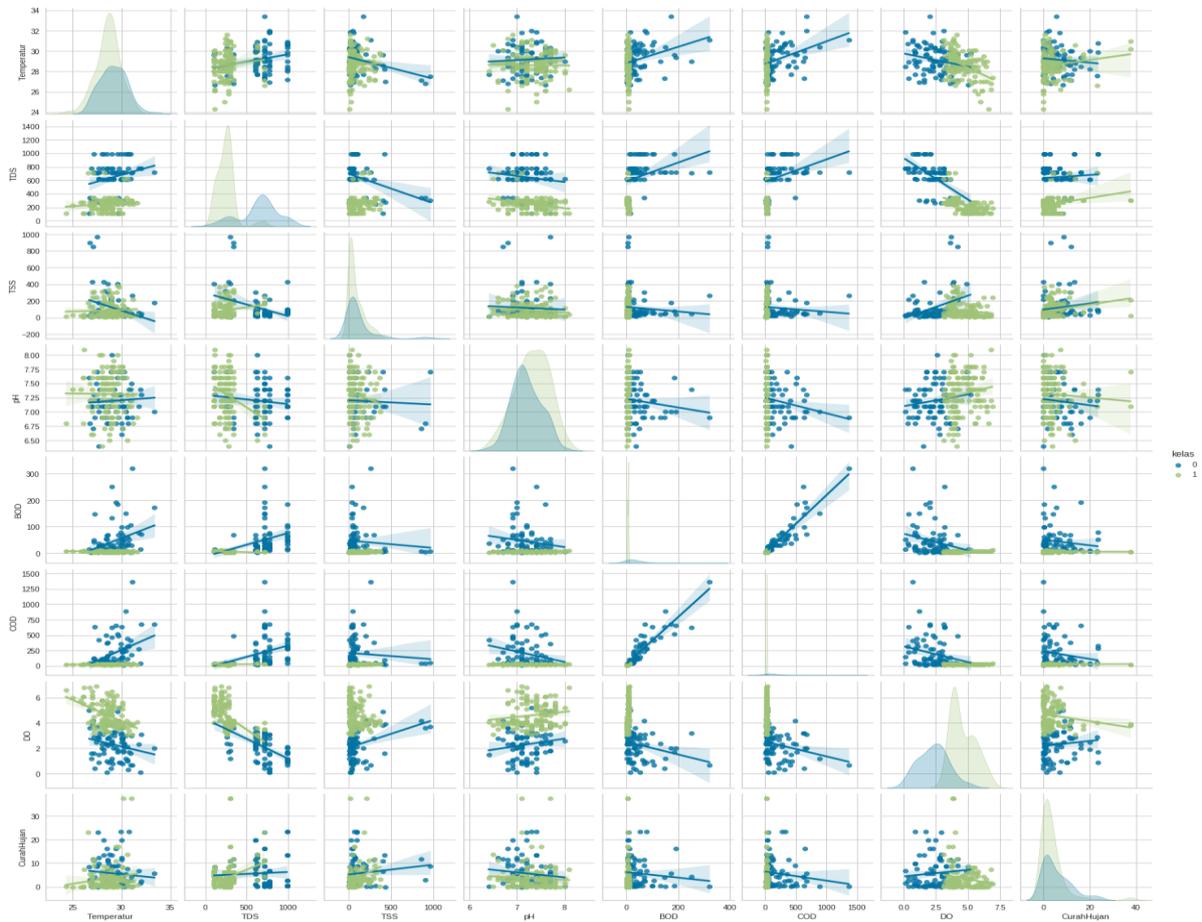
Distribusi data pada 2 label yang ditentukan dapat diketahui kelas 2 dan 3 (peruntukan sarana rekreasi air, perikanan, peternakan, pertanian) yang dilabeli 1 memiliki jumlah data sebesar 61%, sedangkan kelas 4

(peruntukan pertanian) dilabeli 0 memiliki jumlah data sebesar 39%.

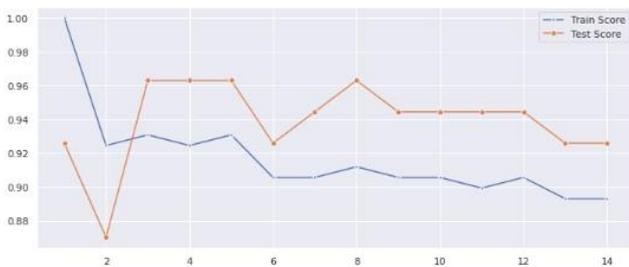
Gambar 3 Menunjukkan perbandingan setiap fitur yang dimiliki dan outlier pada *dataset* pengamatan air sungai dan waduk. Plot berwarna biru gelap merupakan data yang dilabeli 0 yaitu kondisi air yang peruntukannya hanya dapat digunakan sebagai pengelolaan pertanian. Plot berwarna biru muda menunjukkan data yang dilabeli 1 yaitu kondisi air dengan peruntukan sarana rekreasi air (kelas 2), perikanan dan peternakan (kelas 3).



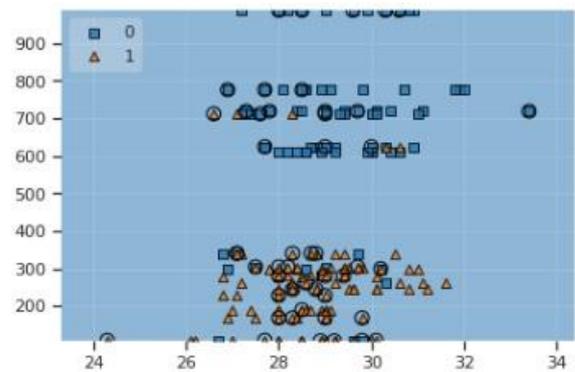
Gambar 4. Distribusi data pada tiap kelas



Gambar 5. Sebaran data berdasarkan perbandingan parameter



Gambar 6. Hasil *training* dan *test* metode *KNN*



Gambar 7. Hasil klasifikasi dengan metode *KNN*

Tabel 3. Perbandingan hasil klasifikasi tunggal dan *ensemble model*

No	Metode	Nilai Akurasi Pada Titik Pengamatan			
		Waduk Sutami Hilir	Cangkir Tambangan	Muara Kali Tengah	Gabungan Titik Pengamatan
1	KNN		Data Tes 33%		
		96	79	96	86
			<i>Cross Validation</i>		
		94	79	89	89
2	SVM		Data Tes 33%		
		95	77	96	87
			<i>Cross Validation</i>		
		94	78	96	89
3	Naïve Bayes		Data Tes 33%		
		94	78	97	90
			<i>Cross Validation</i>		
		93	86	98	91
<i>Ensemble model</i>					
4	Bagging Clasifier				
		92,18	88,90	96,36	94,43
5	Gradient Boosting Clasifier				
		90,18	90,54	96,36	94,96
6	Adaboost Clasifier				
		90,18	90,54	94,54	95,94

4. Kesimpulan

Berdasarkan hasil uji metode klasifikasi tunggal dan *ensemble method* pada data gabungan tiga titik pengamatan diperoleh hasil untuk metode KNN diperoleh akurasi sebesar 86% dengan penggunaan data uji sebesar 33% dari keseluruhan *dataset*. KNN melalui metode *cross validation* (*k-fold validation*) diperoleh nilai akurasi sebesar 89% dengan nilai akurasi tertinggi pada $k=1$ dan $k=4$. Pada penerapan metode SVM nilai c dan jenis kernel yang digunakan mempengaruhi hasil dari akurasi. Nilai akurasi metode SVM 87%, dan 89% untuk penggunaan metode *cross validation*. Metode *Naïve Bayes* memberikan nilai akurasi sebesar 90% dan 91% dengan *cross validation*. Melalui *ensemble model* diperoleh akurasi untuk *Bagging Clasifier* sebesar 94,43%, *Gradient Boosting Clasifier* 94,96%, dan *Adaboost Clasifier* sebesar 95,94%.

Pengamatan pada penerapan metode klasifikasi pada tiga titik pengamatan secara terpisah (seperti yang tercantum dalam tabel 3), diperoleh hasil penurunan dan peningkatan akurasi pada beberapa metode dan titik pengamatan. Ketidakpresisian tersebut disebabkan distribusi data yang semakin tidak seimbang antara label 1 dan 0, serta jumlah *dataset* yang semakin berkurang. Pada penelitian selanjutnya dapat diajukan penerapan *hyperparameter tuning* pada metode klasifikasi *ensemble model* untuk meningkatkan performa dan akurasi dalam melakukan prediksi terhadap data uji baru dari kualitas air sungai dan waduk secara *real time*.

Daftar Pustaka

- [1] S. Yudo, "Kondisi pencemaran logam berat di perairan sungai dki jakarta," *J. Air Indones.*, vol. 2, no. 1, pp. 1–15, 2018, doi: 10.29122/jai.v2i1.2275.
- [2] S. Mariyam, S. Romdon, and E. Kosasih, "Teknik Pengukuran Oksigen Terlarut," *BULETIN TEKNIK LITKAYASA Sumber Daya dan Penangkapan*, vol. 2, no. 1, p. 45, 2017, doi: 10.15578/btl.2.1.2004.45-47.
- [3] K. Aryasa and W. Musu, "Sistem pakar otomatisasi baku mutu limbah pertambangan nikel menggunakan algoritma supervised machine," vol. 2, no. 1, pp. 42–53, 2016.
- [4] A. H. Haghiabi, A. H. Nasrolahi, and A. Parsaie, "Water quality prediction using machine learning methods," *Water Qual. Res. J.*, vol. 53, no. 1, pp. 3–13, 2018, doi: 10.2166/wqrj.2018.025.
- [5] R. Noviandi, Y. H. Chrisnanto, and H. Ashaury, "Sistem segmentasi keluhan air bersih di pt.suryacipta swadaya menggunakan k-medoids clustering," pp. 162–166.
- [6] N. Yulias *et al.*, "Predictin of drinking water facility condition using the naive bayes algorithm," *J. mantik*, vol. 4, no. 4, pp. 2599–2603, 2021.
- [7] T. Wahyono, *Fundamental of python for machine learning: dasar-dasar pemrograman python untuk machine learning dan kecerdasan buatan*, no. September 2018. 2018.
- [8] P. Joshi, *Python machine learning cookBook*. 2016.
- [9] X. Yao and Y. Liu, *Machine learning*. 2005.
- [10] A. P. Wibawa, M. G. A. Purnama, M. F. Akbar, and F. A. Dwiyanto, "Metode-metode klasifikasi," *Pros. Semin. Ilmu Komput. dan Teknol. Inf.*, vol. 3, no. 1, pp. 134–138, 2018.
- [11] D. Xhemali, C. J. Hinde, and R. G. Stone, "Naive bayes vs. decision trees vs. neural networks in the classification of training web pages," *Int. J. Comput. Sci.*, vol. 4, no. 1, pp. 16–23, 2009, [Online]. Available: <http://cogprints.org/6708/>.
- [12] D. Nurdiah and I. A. Muwakhid, "Perbandingan support vector machine dan k-nearest neighbor untuk klasifikasi telur fertil dan infertil berdasarkan analisis texture glm," *J. Transform.*, vol. 13, no. 2, p. 29, 2016, doi: 10.26623/transformatika.v13i2.324.
- [13] A. J. T, D. Yanosma, and K. Anggriani, "Implementasi metode k-nearest neighbor (knn) dan simple additive weighting (saw) dalam pengambilan keputusan seleksi penerimaan anggota paskibraka," vol. III, no. 0065, pp. 98–112, 2016.
- [14] T. G. Dietterich, "Ensemble methods in machine learning," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 1857 LNCS, pp. 1–15, 2000, doi: 10.1007/3-540-45014-9_1.
- [15] R. A. Fitriansyah and Saparudin, "Penerapan ensemble stacking untuk klasifikasi multi kelas," vol. 2, no. 1, pp. 240–243, 2016, [Online]. Available: <http://ars.ilkom.unsri.ac.id>.
- [16] Y. Pristyanto, "Penerapan metode ensemble untuk meningkatkan kinerja algoritme klasifikasi pada imbalanced dataset," *J. Teknoinfo*, vol. 13, no. 1, p. 11, 2019, doi: 10.33365/jti.v13i1.184.
- [17] Peraturan Pemerintah No. 22, "Peraturan pemerintah republik indonesia nomor 22 tahun 2021," *Kementeri. BUMN Indones.*, no. 085459, 2021.
- [18] P. R. Indonesia, *Peraturan pemerintah (pp) tentang penyelenggaraan perlindungan dan pengelolaan lingkungan hidup*, no. 085459. Indonesia: LN.2021/No.32, TLN No.6634, jdih.setkab.go.id : 374 hlm., 2021.
- [19] Afdhaluzzikri, "Analisa kinerja metode naive bayes dengan pembobotan data," 2021.