

Klasifikasi kualitas air dengan menggunakan metode *support vector machine*

Mohamad Arif Abdul Syukur^{1*}, Moh. Heri Susanto², Salman Alfarizhi³

^{1,2,3}Program Studi Teknik Informatika, Universitas Islam Negeri Maulana Malik Ibrahim Malang
e-mail: *200605110044@student.uin-malang.ac.id

Kata Kunci:

kualitas air; SVM;
klasifikasi; lingkungan

Keywords:

water quality; SVM;
classification;
environmental

ABSTRAK

Kualitas air sangat penting bagi kehidupan untuk menjaga keberlanjutan ekosistem lingkungan dalam perairan. Penelitian ini berfokus pada penggunaan Metode Support Vector Machine atau SVM sebagai metode klasifikasi untuk memantau dan mengklasifikasikan kualitas air. Data yang digunakan adalah data indeks kualitas air yang bersumber dari kaggle.com yang berjumlah 8000 data dengan berbagai atribut. Melalui proses pelatihan dan pengujian dengan menggunakan metode SVM didapatkan hasil akurasi mencapai 94,24%. Hasil evaluasi model pada kelas baik untuk nilai presisi adalah 97% dengan recall 91% dan pada kelas

tidak baik nilai presisi adalah 92% dengan recall 98%. Dengan demikian secara keseluruhan model dengan menggunakan metode SVM dapat mengkategorikan kualitas air dengan baik. Sehingga hasil penelitian ini dapat membantu pemerintahan dalam pemantauan kualitas air secara efektif dan lebih cepat terhadap kondisi perairan.

ABSTRACT

Water quality is very important for life to maintain the sustainability of environmental ecosystems in waters. This research focuses on the use of the Support Vector Machine Method or SVM as a classification method for monitoring and classifying water quality. The data used is water quality index data sourced from kaggle.com, amounting to 8000 data with various attributes. Through the training and testing process using the SVM method, accuracy results reached 94.24%. The model evaluation results in the good class for the precision value were 97% with a recall of 91% and in the not good class the precision value was 92% with a recall of 98%. Thus, overall the model using the SVM method can categorize water quality well. So the results of this research can help the government monitor water quality more effectively and more quickly on water conditions.

Pendahuluan

Air merupakan bagian penting dalam kehidupan makhluk hidup karena menjadi salah satu sumber kehidupan. Sekitar 71% wilayah dari bumi merupakan Kumpulan air yang memiliki senyawa kompleks (Hartanti & Ichsan, 2023). Kesehatan Masyarakat dan kelangsungan ekosistem perairan sangat bergantung pada kualitas air. Pertumbuhan populasi manusia dan aktivitas industri dalam beberapa decade terakhir meningkatkan tekanan pada sumber daya air. Sehingga menyebabkan perubahan besar dalam perkembangan kualitas air di seluruh dunia. Air dapat dikatakan bersih dan digunakan untuk kebutuhan sehari-hari jika memenuhi standar yang ditetapkan oleh pemerintahan dan disepakati bersama sesuai kesehatan (Djana, 2023). Kesadaran manusia mengenai



This is an open access article under the [CC BY-NC-SA](https://creativecommons.org/licenses/by-nc-sa/4.0/) license.

Copyright © 2023 by Author. Published by Universitas Islam Negeri Maulana Malik Ibrahim Malang.

pola konsumsi air berkualitas semakin tinggi sehingga menghasruskan pemantauan kualitas air pada setiap wilayah perairan (Ulum et al., 2023). Oleh karena itu pemantauan kualitas air sangat penting dalam mengetahui dampak aktivitas manusia terhadap lingkungan perairan. Tujuan penelitian ini adalah untuk mengembangkan metode klasifikasi dengan menggunakan SVM dalam mengklasifikasi kualitas air pada data yang didapatkan dari Kaggle.com.

Support vector machine adalah metode yang dibuat oleh Vapnik pada tahun 1995. Seiring berjalannya waktu banyak perkembangan hingga sampai saat ini menjadi salah satu metode klasifikasi yang banyak digunakan untuk menangani permasalahan pada dataset yang kompleks dan tidak linear. Sehingga pada penelitian ini SVM akan digunakan sebagai metode untuk mengklasifikasikan kualitas air pada dataset yang didapatkan dari Kaggle.com dengan atribut yang memeberikan informasi tentang kesehatan air dan kondidi ekologis. Support vector machine merupakan sistem pembelajaran yang dilatih dengan berdasarkan teori optimasi menggunakan ruang hipotesis yang berupa fungsi linear dalam dimensi tinggi (Putri et al., 2023).

Penelitian sebelumnya tentang pemantauan kualitas air telah banyak memanfaatkan teknologi informasi dan metode komputasi. Namun sebagian besar penelitian tersebut berkonsentrasi pada analisis deskriptif dan kurang mempelajari metode klasifikasi yang digunakan seperti metod SVM untuk meningkatkan ketepatan dan efisisensi dalam mengklasifikasikan kualitas air. Kualitas air dapat diklasifikasikan dengan menggunakan metode SVM karen metode ini mempunyai banyak kelebihan. Metode ini bekerja dengan baik pada data ynag berdimensi tinggi karena sudah banyak dikembangkan oleh beberapa ilmuan untuk data yang nonlinear. Cara kerja SVM adalah dnegan mencari garis hyperplane dan margin untuk memisahkan data sehingga pemisahan klasifikasi akan lebih maksimal (Sudin et al., 2023).

Pemantauan kualitas air sangat penting untuk ekosistem dan kesehatan manusia serta lingkungan sekitar karena dapat menimbulkan pencemaran air yang tidak baik. Sehingga perlu adanya pemantauan kualitas air. Dalam hal ini metode SVM dapat digunakan untuk klasifikasi kualitas air dengan baik untuk menemukan pola dan hubungan yang kompleks antara berbagai kelas pada data yang digunakan. Sehingga dapat mendapatkan informasi tentang kualitas air dengan lebih cepat dan dapat digunakan untuk membuat Keputusan pengelolaan sumber daya air yang lebih baik. Dengan demikian untuk memenuhi standar kualitas air akan lebih mudah karena terdapat adanya upaya pengawasan atau pemantauan sumber air. Sehingga air akan terjaga kualitasnya dan mampu menghasilkan standar air yang layak dikonsumsi (Savitri & Nursalim, 2023).

Berdasarkan latar belakang tersebut, pada penelitian ini dirancang dan dibangun sebuah model untuk sistem klasifikasi kualitas air pada data yang didapatkan dari Kaggle. Penelitian ini berfokus pada pengembangan metode SVM untuk mengatasi pengklasifikasian pada kelas target. Karena SVM adalah metode yang mendalam dan praktis dalam pengklasifikasian data dan terbukti efektif dalam berbagai konteks klasifikasi. Penelitian ini diharapkan dapat memberikan pemahaman yang lebih baik tentang kualitas air dan membantu mempermudah keputusan yang lebih cepat dalam pengelolaan sumber daya air yang berkelanjutan. Penelitian ini diharapkan dapat

menjadi solusi baru untuk permasalahan dalam penanganan kualitas air. Dengan menggunakan metode SVM juga diharapkan memberi peneliti dan praktisi ataupun pengambil kebijakan keputusan seperti pemerintahan dalam panduan untuk mengelola dan menjaga kualitas air di lingkungan pedesaan ataupun perkotaan.

Rumusan masalah yang menjadi pokok bahasan penelitian ini adalah bagaimana merancang dan membangun model untuk sistem klasifikasi kualitas air dengan menggunakan metode SVM. Sehingga tujuan dari penelitian ini adalah merancang dan membangun model untuk sistem kualitas air dengan menggunakan metode SVM.

Metode

Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah salah satu metode algoritma pembelajaran mesin yang paling kuat dan populer digunakan untuk klasifikasi dan regresi. Metode ini pertama kali diperkenalkan oleh Vladimir Vapnik dan koleganya di AT&T Bell Laboratories pada tahun 1992. SVM bekerja dengan menemukan hyperplane yang optimal sehingga dapat memisahkan kelas dalam dataset dengan menggunakan margin yang besar. Konsep utama SVM adalah memaksimalkan margin yang mana jarak antara hyperplane dan titik data terdekat dari kedua kelas yang dikenal dengan support vectors (Angellina et al., 2023).

Pada metode SVM dua dimensi, hyperplane dapat dianggap sebagai garis yang membagi data ke dalam dua kelas yang berbeda. Rumus dasar untuk hyperplane adalah sebagai berikut :

$$w \cdot x - b = 0 \quad (1)$$

w = vector bobot

x = vector fitur

b = bias

Tujuan SVM adalah menemukan w dan b yang memaksimalkan margin antara dua kelas data. Titik-titik data yang berada paling dekat dengan hyperplane disebut vectors dan titik-titik itulah yang menjadi kunci dalam menentukan posisi hyperplane.

Untuk kasus data yang tidak sepenuhnya dapat dipisahkan secara linear maka SVM dikembangkan dengan menggunakan kernel trick dengan tujuan untuk memetakan data ke ruang dimensi yang lebih tinggi agar data dapat dipisahkan secara linear. Terdapat beberapa kernel pada SVM, kernel yang umum digunakan termasuk polynomial, radial basis function atau RBF dan sigmoid (Hermawan, 2019). Misalkan kernel RBF berikut adalah rumusnya :

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (2)$$

$K(x_i, x_j)$ = kernel

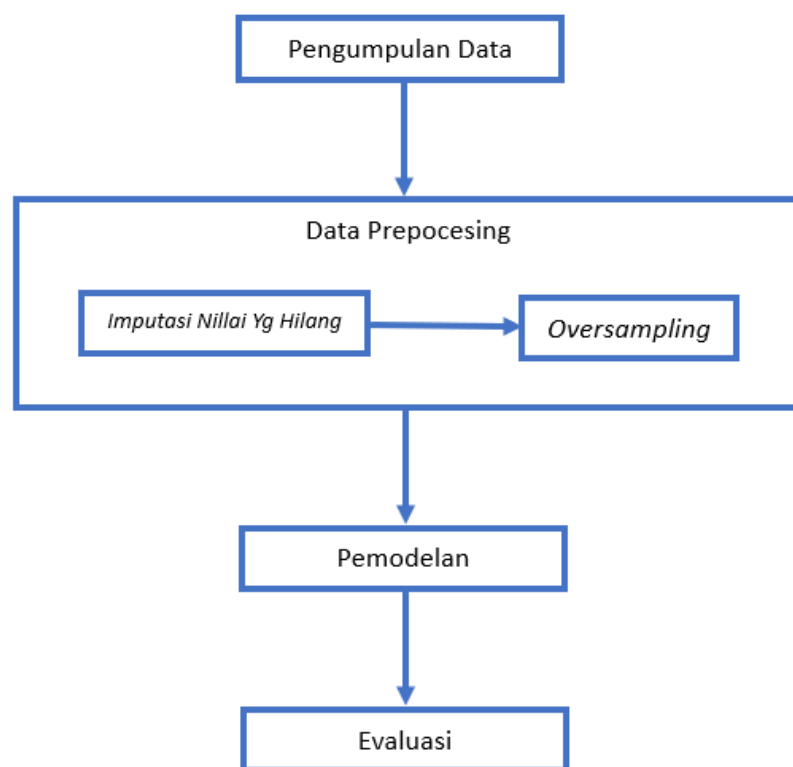
γ = parameter yang menentukan jangkauan pengaruh dari satu titik data

Fungsi kernel adalah menggantikan produk dot dalam ruang fitur asli yang memungkinkan SVM untuk menangani klasifikasi non-linear.

SVM adalah algoritma yang sangat kuat karena beberapa kelebihanannya yaitu mampu bekerja dengan baik pada dataset dengan dimensi tinggi dan efektif dalam kasus jumlah fitur yang lebih besar dari jumlah sampel. SVM juga tidak menggunakan memori yang banyak sehingga efisien karena hanya menggunakan subset dari titik pelatihan atau support vectors dalam fungsi keputusan (Jayadi et al., 2023).

Pembahasan

Penelitian yang akan dikembangkan menggunakan metode SVM diperlukan suatu tahapan yang harus dilakukan berikut adalah tahapan-tahapan yang akan dilakukan pada penelitian ini.



Gambar 1. Rancangan Sistem.

Pengumpulan Data

Pengumpulan data adalah tahap awal dalam suatu penelitian yang merupakan proses mengumpulkan informasi atau data dari berbagai sumber untuk di analisis dan dijadikan bahan penelitian. Dalam penelitian ini data yang akan digunakan diambil dari Kaggle.com yang merupakan data public. Kaggle sendiri adalah web yang menyediakan berbagai jenis dataset berupa file csv, excel, json ataupun format lainnya. Dataset untuk penelitian ini berjumlah 8000 data kualitas air di lingkungan perkotaan dengan beberapa atribut lengkap.

Dataset yang dikumpulkan adalah data waterQuality.csv yang bersumber dari kaggle. Pada dataset tersebut terdapat 21 atribut yaitu aluminium, ammonia, arsenic,

barium, cadmium, chloramine, chromium, copper, flouride, bacteria, viruses, lead, nitrates, nitrites, mercury, perchlorate, radium, selenium, silver, uranium, dan terakhir adalah is_safe yaitu kelas atribut dengan rating 0 dan 1. Jika 0 maka kualitas air dikatakan tidak sehat dan jika 1 maka kualitas air adalah sehat. Berikut adalah data yang digunakan.

	aluminium	ammonia	arsenic	barium	cadmium	chloramine	chromium	copper	flouride	bacteria	...	lead	nitrates	nitrites	mercury	perchlorate	radium
0	1.65	9.08	0.04	2.85	0.007	0.35	0.83	0.17	0.05	0.20	...	0.054	16.08	1.13	0.007	37.75	6.78
1	2.32	21.16	0.01	3.31	0.002	5.28	0.68	0.66	0.90	0.65	...	0.100	2.01	1.93	0.003	32.26	3.21
2	1.01	14.02	0.04	0.58	0.008	4.24	0.53	0.02	0.99	0.05	...	0.078	14.16	1.11	0.006	50.28	7.07
3	1.36	11.33	0.04	2.96	0.001	7.23	0.03	1.66	1.08	0.71	...	0.016	1.41	1.29	0.004	9.12	1.72
4	0.92	24.33	0.03	0.20	0.006	2.67	0.69	0.57	0.61	0.13	...	0.117	6.74	1.11	0.003	16.90	2.41
...
7994	0.05	7.78	0.00	1.95	0.040	0.10	0.03	0.03	1.37	0.00	...	0.197	14.29	1.00	0.005	3.57	2.13
7995	0.05	24.22	0.02	0.59	0.010	0.45	0.02	0.02	1.48	0.00	...	0.031	10.27	1.00	0.001	1.48	1.11
7996	0.09	6.85	0.00	0.61	0.030	0.05	0.05	0.02	0.91	0.00	...	0.182	15.92	1.00	0.000	1.35	4.84
7997	0.01	10	0.01	2.00	0.000	2.00	0.00	0.09	0.00	0.00	...	0.000	0.00	0.00	0.000	0.00	0.00
7998	0.04	6.85	0.01	0.70	0.030	0.05	0.01	0.03	1.00	0.00	...	0.182	15.92	1.00	0.000	1.35	4.84

7999 rows x 21 columns

Gambar 2. Dataset.

Imputasi Data Yang Hilang

Imputasi data yang hilang adalah langkah yang penting dalam penelitian karena berguna untuk merapihkan dataset sebelum digunakan untuk membangun model. Proses ini memastikan bahwa data yang hilang diisi secara akurat agar data menjadi valid dan tidak ada masalah serta bisa melakukan analisis data lebih lanjut (Jasman et al., 2022). Data yang hilang biasanya ditandai dengan nilai NaN (Not a Number) atau tanda lain seperti #NUM!, NA, atau null. Pada dataset yang digunakan terdapat data yang hilang yang didefinisikan dengan nilai #NUM!. Sehingga untuk memproses data lebih lanjut, data tersebut harus diubah dan diisi dengan baik sesuai type data pada atribut. Berikut adalah data sebelum dan sesudah diimputasi.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7999 entries, 0 to 7998
Data columns (total 21 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   aluminium       7999 non-null   float64
1   ammonia         7999 non-null   object
2   arsenic         7999 non-null   float64
3   barium          7999 non-null   float64
4   cadmium         7999 non-null   float64
5   chloramine      7999 non-null   float64
6   chromium        7999 non-null   float64
7   copper          7999 non-null   float64
8   flouride        7999 non-null   float64
9   bacteria        7999 non-null   float64
10  viruses         7999 non-null   float64
11  lead            7999 non-null   float64
12  nitrates        7999 non-null   float64
13  nitrites        7999 non-null   float64
14  mercury         7999 non-null   float64
15  perchlorate     7999 non-null   float64
16  radium          7999 non-null   float64
17  selenium        7999 non-null   float64
18  silver          7999 non-null   float64
19  uranium         7999 non-null   float64
20  is_safe         7999 non-null   object
dtypes: float64(19), object(2)
memory usage: 1.3+ MB
```

Gambar 1. Data Sebelum Imputasi

```

Kolom 'ammonia' memiliki data bukan numerik:
7551      #NUM!
7568      #NUM!
7890      #NUM!
Name: ammonia, dtype: object
Kolom 'is_safe' memiliki data bukan numerik:
7551      #NUM!
7568      #NUM!
7890      #NUM!
Name: is_safe, dtype: object

```

Gambar 2. Pengecekan data non-numerik

```

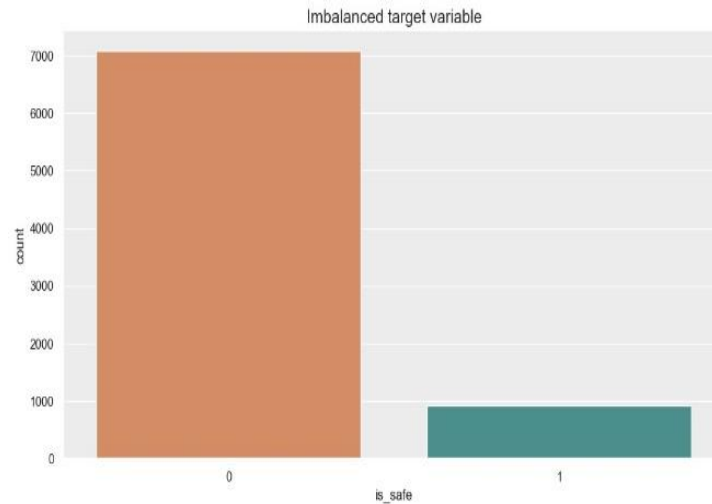
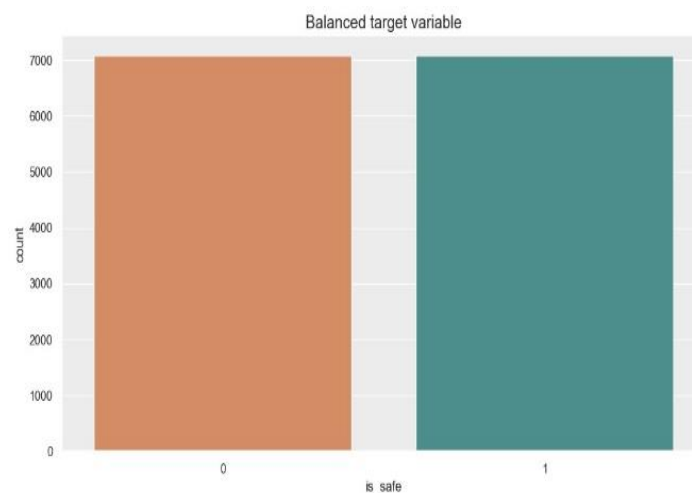
<class 'pandas.core.frame.DataFrame'>
Index: 7996 entries, 0 to 7998
Data columns (total 21 columns):
#   Column          Non-Null Count  Dtype
---  -
0   aluminium       7996 non-null   float64
1   ammonia         7996 non-null   float64
2   arsenic         7996 non-null   float64
3   barium          7996 non-null   float64
4   cadmium         7996 non-null   float64
5   chloramine      7996 non-null   float64
6   chromium        7996 non-null   float64
7   copper          7996 non-null   float64
8   flouride        7996 non-null   float64
9   bacteria        7996 non-null   float64
10  viruses         7996 non-null   float64
11  lead            7996 non-null   float64
12  nitrates        7996 non-null   float64
13  nitrites        7996 non-null   float64
14  mercury         7996 non-null   float64
15  perchlorate     7996 non-null   float64
16  radium          7996 non-null   float64
17  selenium        7996 non-null   float64
18  silver          7996 non-null   float64
19  uranium         7996 non-null   float64
20  is_safe         7996 non-null   int64
dtypes: float64(20), int64(1)
memory usage: 1.3 MB

```

Gambar 3. Data Sesudah Imputasi

Oversampling

Oversampling adalah suatu teknik dalam pemrosesan data yang digunakan untuk menangani kelas yang tidak seimbang menjadi seimbang (Piyadasa & Gunawardana, 2023). Ketidakseimbangan kelas biasanya terjadi ketika terdapat perbedaan jumlah data pada masing masing kelas (Afrah et al., 2024). Setelah dilakukan imputasi pada tahap sebelumnya maka tahap selanjutnya adalah proses oversampling yang diharapkan mendapatkan data yang seimbang agar dataset pada proses selanjutnya dalam pembangunan model akan lebih baik. Berikut adalah gambar data sebelum proses oversampling dan sesudah proses oversampling.

**Gambar 4.** Data Imbalanced**Gambar 5.** Data Balanced

Splitting Data

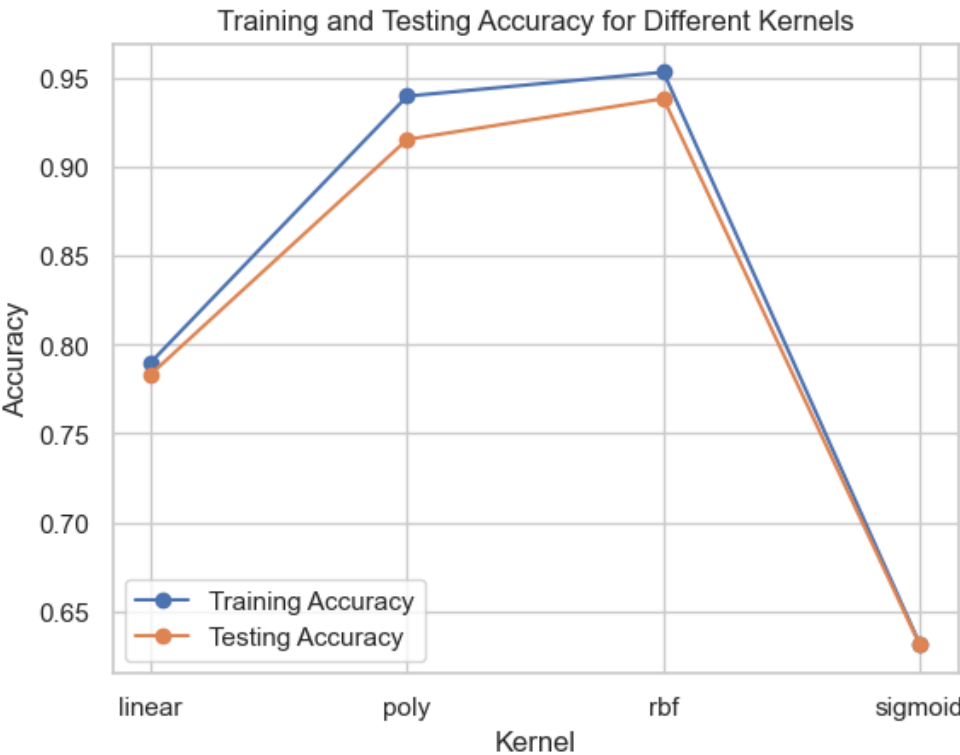
Splitting data adalah proses pemisahan menjadi dua bagian yaitu training dan testing (Sari et al., 2023). Data tersebut akan digunakan untuk mengukur model. Data training berfungsi untuk pelatihan data yang digunakan pada saat pembuatan model. sedangkan data testing adalah untuk digunakan sebagai pengujian data yang akan diuji dari hasil prediksi yang dibuat berdasarkan data training (Okprana & Winanjaya, 2022). Pada proses ini data yang sudah seimbang dari hasil oversampling tahap sebelumnya akan dibagi menjadi dua bagian yaitu data training sebesar 80% dan data testing 20%. Dengan demikian 80% data training akan digunakan untuk mengukur nilai akurasi pelatihan dengan jumlah 6400 data. Sedangkan sisanya adalah data testing yang akan digunakan untuk pengujian model SVM berdasarkan model training yang sudah didapatkan.

Tabel 1. Splitting Data.

Training	Testing
80%	20%

Pembangunan Model SVM

Support Vector Machine adalah algoritma yang efektif dalam melakukan klasifikasi. Pada penelitian ini kualitas air akan diprediksi menggunakan SVM berdasarkan berbagai atribut yang diukur dari air sesuai dengan atribut yang ada pada dataset. Sehingga pada tahap ini SVM akan mencari pola atau batasan yang memisahkan berbagai kategori kualitas air melalui proses training atau pelatihan. Tujuan utamanya adalah untuk membangun model yang dapat membedakan antara kategori kualitas air yang baik dan buruk dengan tingkat akurasi yang tinggi. Dalam pemodelan menggunakan SVM terdapat beberapa jenis kernel yang digunakan pada saat pengujian untuk menisolasi data. Namun melalui serangkaian uji coba dan eksperimen kernel Radial Basis Function atau RBS adalah uji coba dengan hasil yang paling tinggi dibandingkan dnegan kernel lainnya. Sehingga memungkinkan SVM menangani data non linear dengan baik dibandingkan kernel lain. Berikut adalah hasil dari pengujian masing-masing kernel.



Gambar 6. Visualisasi Akurasi

Evaluasi

Setelah Pembangunan model dilakukan tahap selanjutnya adalah evaluasi model pada data uji untuk mengetahui seberapa baik kinerja model yang sudah dibuat. Seberapa baik model dapat mengklasifikasikan kelas kualitas air berdasarkan atribut

yang diberikan dapat dinilai dengan menggunakan metrik evaluasi seperti akurasi, presisi, recall dan f1-score (Sahi et al., 2023).

Akurasi menggambarkan besaran data yang diklasifikasikan sebagai benar dibandingkan keseluruhan data.

$$\text{Akurasi} = (\text{Jumlah prediksi benar}) / (\text{Total Data}) \times 100\%$$

Sementara presisi, menyoroti seberapa tepat model dalam mengidentifikasi contoh positif.

$$\text{Presisi} = \text{TP} / ((\text{TP} + \text{FP}))$$

True Positive (TP) : Jumlah hasil identifikasi positif yang benar.

False Positive (FP) : Jumlah hasil identifikasi negatif yang keliru.

Sedangkan recall, menunjukkan kemampuan model untuk menemukan semua contoh positif.

$$\text{Recall} = \text{TP} / ((\text{TP} + \text{FN}))$$

False Negatif (FN) : Jumlah positif yang seharusnya teridentifikasi, tetapi tidak teridentifikasi.

Dan F1-Score, sebagai pengukuran gabungan dari presisi dan recall, memberikan gambaran holistik tentang kinerja model.

$$\text{F1-Score} = 2 \times \text{Precision} \times \text{Recall} / (\text{Precision} + \text{Recall})$$

Dalam prediksi kualitas air SVM dapat digunakan untuk membuat prediksi yang berguna berdasarkan atribut yang diukur. Hal ini dapat mendukung upaya pemantauan dan perlindungan lingkungan terhadap sumber air dengan memberikan informasi tentang tingkat kebersihan atau kesesuaian air untuk penggunaan tertentu. Dari percobaan yang dilakukan dalam prediksi kualitas air dengan menggunakan SVM menunjukkan hasil yang sangat baik mencapai 94,24%. Dengan demikian sebagian besar data yang digunakan merupakan data dengan kualitas air yang baik. Dan juga bisa dikatakan bahwa data yang digunakan cocok dengan model SVM yang digunakan. Berikut adalah gambar hasil akurasi pengujian.

Classification Report:				
	precision	recall	f1-score	support
0	0.98	0.90	0.94	1390
1	0.91	0.98	0.95	1444
accuracy			0.94	2834
macro avg	0.95	0.94	0.94	2834
weighted avg	0.94	0.94	0.94	2834

Gambar 7. Hasil Akurasi

Pada Gambar 9 diatas terlihat bahwa nilai precision untuk kelas 0 menunjukkan nilai 98% sehingga menunjukkan bahwa sebagian besar diklasifikasikan sebagai kelas 0 dan data sebenarnya memang benar dari kelas 0. Sementara untuk precision dikelas 1

menunjukkan nilai 91% dengan demikian berarti menunjukkan bahwa sebagian kecil diklasifikasikan dengan salah. Precision dapat dikatakan baik apabila memiliki nilai mendekati 1. Berikut adalah gambar confusion matrix yang digunakan untuk evaluasi model yang dilakukan.



Gambar 8. Confusion Matrix

Terlihat bahwa nilai data kualitas air yang tidak aman berjumlah 1260 karena sesuai dengan data yang sebenarnya pada dataset yang digunakan dan sesuai dengan prediksi dari metode SVM. Sedangkan yang diprediksi aman akan tetapi pada data yang sebenarnya tidak aman sebesar 130 data. Untuk data kualitas air yang aman diprediksi aman oleh system dan pada data yang sebenarnya juga aman berjumlah 1411 data sedangkan yang diprediksi aman akan tetapi pada data yang sebenarnya adalah tidak aman berjumlah 33 data.

Kesimpulan dan Saran

Dari penelitian yang sudah dilakukan dalam pengklasifikasian kualitas air menggunakan metode SVM didapatkan kesimpulan bahwa metode SVM menghasilkan akurasi sebesar 94%, precision sebesar 95%, recall sebesar 94%, dan f1-score sebesar 94% dengan jumlah data yang digunakan adalah 8000 data dengan 21 atribut. Berdasarkan hasil yang didapatkan tersebut maka dapat ditarik kesimpulan bahwa metode SVM dapat digunakan untuk pengklasifikasian kualitas air dengan kinerja yang sangat baik.

Daftar Pustaka

Afrah, A. S., Sari, N. F. A. T., Utama, S. N., Holle, K. F. H., Lestandy, M., Sintiya, E. S., & Rizdania. (2024). Comparative Study of Machine learning and holt-winters exponential smoothing models for prediction of CPI's seasonal data. *IEEE Xplore*, 144–148. <https://doi.org/10.1109/ICoSEIT60086.2024.10497509>

- Angellina, Herwindiati, D. E., & Hendryli, J. (2023). Performa support vector machine pada klasifikasi lahan dan air tanah. *Jurnal Media Informatika Budidarma*, 231–241. <https://doi.org/10.30865/mib.v7i1.5279>
- Djana, M. (2023). Analisis kualitas air dalam pemenuhan kebutuhan air bersih di Kecamatan Natar Hajimena Lampung Selatan. *Jurnal Redoks*, 8(1), 81–87. <https://doi.org/10.31851/redoks.v8i1.11853>
- Hartanti, D., & Ichsan, A. (2023). Komparasi algoritma machine learning dalam identifikasi kualitas air. *SMARTICS Journal*, 9(1), 1–6. <https://doi.org/10.21067/smartics.v9i1.8113>
- Hermawan, A. (2019). *SPKU: Sistem Prediksi Kualitas Udara (Studi Kasus: DKI Jakarta)*. (Tugas Akhir Naskah Publik, Universitas Teknologi Yogyakarta). <http://eprints.uty.ac.id/3552/>
- Jasman, T. Z., Fadhlullah, M. A., Pratama, A. L., & Rismayani, R. (2022). Analisis algoritma gradient boosting, adaboost dan catboost dalam klasifikasi kualitas air. *JuTISI: Jurnal Teknik Informatika dan Sistem Informasi*, 8(2), 392–402. <https://doi.org/10.28932/jutisi.v8i2.4906>
- Jayadi, B. V., Handhayani, T., & Lauro, M. D. (2023). Perbandingan KNN dan SVM untuk klasifikasi kualitas udara di Jakarta. *JIKSI: Jurnal Ilmu Komputer dan Sistem Informasi*, 11(1), 1–7. <https://doi.org/10.24912/jiksi.v11i2.26006>
- Okprana, H., & Winanjaya, R. (2022). Analisis Pengaruh komposisi data training dan testing terhadap akurasi algoritma Resilient Backpropagation (RProp). *Brahmana: Jurnal Penerapan Kecerdasan Buatan*, 4(1), 89–95. <https://tunasbangsa.ac.id/pkm/index.php/brahmana/article/view/138>
- Piyadasa, T. D., & Gunawardana, K. (2023). A review on oversampling techniques for solving the data imbalance problem in classification. *ICTer: International Journal on Advances in ICT for Emerging Regions*, 16(1), 22–31. <https://doi.org/10.4038/icter.v16i1.7260>
- Putri, S. S. M., Arhami, M., & Hendrawaty. (2023). Penerapan metode SVM pada klasifikasi kualitas air. *JAISE: Journal of Artificial Intelligence and Software Engineering*, 3(2), 94–101. <http://dx.doi.org/10.30811/jaise.v3i2.4630>
- Sahi, M., Faisal, M., Arif, Y. M., & Crysdian, C. (2023). Analysis of the use of artificial neural network models in predicting Bitcoin prices. *AIMS: Applied Information System and Management*, 6(2), 91–96. <https://doi.org/10.15408/aism.v6i2.29648>
- Sari, N. F. A. T., Nabela, M., & Abdurrohman, M. F. (2023). Utilizing the K-means algorithm for breast cancer diagnosis: A promising approach for improved early detection. *MATICS: Jurnal Ilmu Komputer dan Teknologi Informasi (Journal of Computer Science and Information Technology)*, 15(2), 72–78. <https://doi.org/10.18860/mat.v15i2.23644>
- Savitri, L., & Nursalim, R. (2023). Klasifikasi kualitas air minum menggunakan penerapan algoritma machine learning dengan pendekatan supervised learning. *Diophantine Journal of Mathematics and Its Applications*, 2(1), 30–36. <https://doi.org/10.33369/diophantine.v2i01.28260>
- Sudin, A., Salmin, M., Fhadli, M., & Mamonto, A. M. (2023). Klasifikasi kelayakan air minum bagi tubuh manusia menggunakan metode support vektor machine dengan backward elimination. *Jurnal Jaringan dan Teknologi Informasi*, 2(1), 87–95. <https://e->

journal.unkhair.ac.id/index.php/jati/article/view/61

Ulum, S., Alifa, R. F., Rizkika, P., & Rozikin, C. (2023). Perbandingan performa algoritma KNN dan SVM dalam klasifikasi kelayakan air minum. *Generation Journal*, 7(2), 141–146. <https://doi.org/10.29407/gj.v7i2.20270>