

Klasifikasi dengan metode *Support Vector Machine* (SVM) dalam menentukan kualitas air minum

Nenden Nuraeni

Program Studi Teknik Informatika, Universitas Islam Negeri Maulana Malik Ibrahim Malang
e-mail: 210605110149@student.uin-malang.ac.id

Kata Kunci:

Kualitas air, evaluasi, kernel, klasifikasi, SVM

Keywords:

Water quality, evaluation, kernel, classification, SVM

ABSTRAK

Ketersediaan air bersih yang semakin terbatas menjadi masalah global yang serius. Krisis air bersih berdampak signifikan pada kesehatan global, termasuk kematian akibat air terkontaminasi. Proses pengujian air yang memakan waktu lama menjadi hambatan utama dalam menentukan kualitas air. Penelitian ini menggunakan metode Support Vector Machine (SVM) untuk klasifikasi kualitas air. Data yang digunakan berasal dari kumpulan berbagai senyawa dan unsur yang digunakan dalam pengujian kualitas air. Penelitian ini melibatkan tahapan pengumpulan data, pembersihan data, pelabelan data, dan pembagian data menjadi data latih dan data uji dengan tiga skenario pengujian. Evaluasi model dilakukan dengan mengukur akurasi, presisi, dan recall. Hasil penelitian menunjukkan bahwa SVM dengan kernel RBF memberikan performa terbaik dalam memprediksi kualitas air minum dengan akurasi sebesar 95%. Temuan ini berkontribusi penting dalam pengembangan metode pengujian kualitas air yang lebih akurat dan efisien.

ABSTRACT

The increasingly limited availability of clean water is becoming a serious global problem. The clean water crisis significantly impacts global health, including deaths from contaminated water. The time-consuming process of water testing is a significant obstacle in determining water quality. This research uses the Support Vector Machine (SVM) method for water quality classification. The data used comes from a collection of various compounds and elements used in water quality testing. This research involves the stages of data collection, data cleaning, data labeling, and data division into training data and test data with three test scenarios. Model evaluation was done by measuring accuracy, precision, and recall. The results showed that SVM with RBF kernel best predicted drinking water quality with 95% accuracy. These findings contribute to developing more accurate and efficient water quality testing methods.

Pendahuluan

Air adalah senyawa penting yang mendukung kehidupan semua makhluk di bumi. Sebagai pelarut universal, air memungkinkan reaksi-reaksi kimia dalam sel dan organisme, menyusun sekitar 60-70% massa sel, serta berfungsi dalam transportasi nutrisi dan oksigen, dan mengatur suhu tubuh (Tatarov, 2022). Namun, krisis air bersih semakin mengkhawatirkan. Menurut WHO hanya sekitar 1% dari total air di planet ini yang bisa dikonsumsi manusia, sehingga menyebabkan banyak orang kesulitan mengakses air minum yang layak (Koninck et al., 2023). Menurut WHO, 663 juta orang di dunia masih kesulitan mendapatkan air bersih, dan diperkirakan pada 2025, sekitar



This is an open access article under the CC BY-NC-SA license.

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

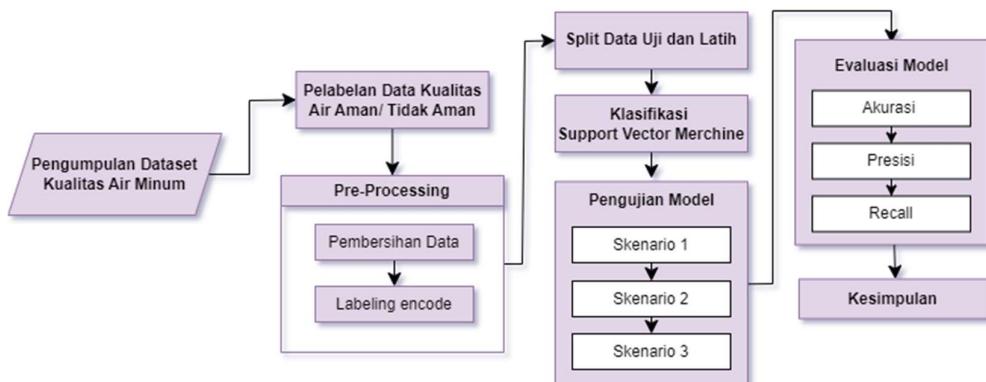
dua pertiga populasi dunia akan tinggal di daerah dengan kelangkaan air (Mishra & Tushaus, 2022; Shemer et al., 2023). Setiap tahun, 829 ribu orang meninggal akibat air minum yang tidak aman, sanitasi buruk, dan kebersihan tangan yang tidak memadai (Shemer et al., 2023). Prediksi WWAP menyebutkan 85% air bersih yang tersedia akan menjadi limbah, sementara kebutuhan air terus meningkat (Zhou et al., 2019). Krisis ini mengancam kesehatan dan kesejahteraan global, menyebarkan penyakit dari air terkontaminasi yang bisa mematikan, terutama bagi anak-anak, serta memengaruhi kebersihan dan sanitasi (Alijanzadeh Maliji et al., 2023; Devgade & Patil, 2023). Dalam menentukan kualitas air, diperlukan waktu yang cukup lama karena harus melalui serangkaian pengujian parameter fisik, kimia, dan biologi (Dhaswadikar Usha Sitaram, 2022; Kashtanjeva et al., 2022). Meningkatnya kesadaran masyarakat terhadap kualitas air mendorong kebutuhan akan metode yang dapat memastikan air yang dikonsumsi atau digunakan aman dan bermutu. Oleh karena itu, diperlukan teknologi dan metode pengujian dengan akurasi tinggi untuk memastikan bahwa air yang dikonsumsi benar-benar aman dan berkualitas.

Penelitian mengenai kualitas air telah menjadi fokus utama dalam beberapa studi terkini. Sebuah penelitian yang dilakukan oleh Tanega dkk menggunakan metode Random Forest untuk mengklasifikasikan kualitas air Danau Tall di Filipina, menghasilkan tingkat akurasi yang sangat tinggi sebesar 95% (Tanega et al., 2023). Penelitian serupa oleh Adi T dkk juga dilakukan dengan menerapkan tiga metode klasifikasi data yang berbeda, dan menghasilkan akurasi tertinggi masing-masing sebesar 71,19% untuk K-nearest neighbours, 67,16% untuk Naïve Bayes, dan 61,77% untuk Decision Tree (Tangkelayuk, 2022). Namun, penelitian yang dilakukan oleh (Marten Sahalatua Tumangger & Hidayat, 2019), menggunakan metode SVM dengan kernel linier, menghasilkan tingkat akurasi yang relatif rendah sebesar 78,70%, berbeda dengan penelitian lain yang dilakukan oleh mardewi dkk yang menerapkan metode SVM dengan tiga kernel (linear, polynomial, RBF), mampu menghasilkan tingkat akurasi yang jauh lebih tinggi, yakni mencapai 95,0% untuk kernel linear dan polynomial, serta 94,5% untuk RBF (Mardewi et al., 2023). Temuan ini menyoroti pentingnya variasi dalam penggunaan kernel pada metode SVM dalam konteks klasifikasi kualitas air, dengan implikasi signifikan bagi penelitian selanjutnya dalam domain ini.

Penggunaan metode seperti K-nearest neighbours, Naïve Bayes, dan Decision Tree memiliki kelemahan dalam menangani data yang kompleks serta cenderung menghasilkan kesalahan saat terdapat ketidakseimbangan dalam jumlah sampel pada setiap kelas (Aldossary et al., 2022). Demikian pula, meskipun Random Forest menunjukkan tingkat akurasi yang tinggi, namun memiliki batasan dalam menangani masalah kompleksitas dan tidak linearan pada data kualitas air. Sebaliknya, Support Vector Machine (SVM) terbukti lebih efektif dalam menangani tantangan tersebut, terutama pada data yang memiliki struktur yang rumit atau ketidakseimbangan antar kelas (Santoso et al., 2022). Oleh karena itu, penelitian ini mengusulkan penggunaan metode SVM untuk memprediksi kualitas air minum. Metode SVM yang diusulkan akan menggunakan empat kernel dan melibatkan tiga skema pengujian. Model dengan akurasi tertinggi dari kernel yang dihasilkan akan dipilih untuk memprediksi kualitas air.

Metode Penelitian

Klasifikasi kualitas air minum merupakan proses penting untuk memastikan bahwa air yang dikonsumsi memenuhi standar keselamatan dan kesehatan. Metode Support Vector Machine (SVM) digunakan dalam penelitian ini untuk mengklasifikasikan kualitas air minum dengan berbagai jenis kernel, seperti sigmoid, linear, RBF (Radial Basis Function), dan polynomial. Proses ini melibatkan beberapa tahapan yang dijelaskan dan digambarkan dalam sebuah diagram alir pada Gambar 1.



Gambar 1. Kerangka Penelitian

Pengumpulan Data

Data yang digunakan dalam penelitian klasifikasi kualitas air minum ini berasal dari kumpulan berbagai senyawa dan unsur yang digunakan dalam pengujian kualitas air, yang terdiri dari 21 atribut dan 8000 data. Atribut-atribut tersebut dijabarkan dalam Tabel 1.

Tabel 1. Data Set

Aluminum	Ammonia	...	Uranium	Is_Safe
0,086806	09.08	...	00.02	1
02.32	21.16	...	00.05	1
01.01	14.02	...	00.01	0

Sumber: <https://www.kaggle.com/datasets/mssmartypants/water-quality>

Pelabelan Data

Pelabelan data merupakan proses penting dalam analisis dan pengolahan data, terutama dalam konteks klasifikasi dan pembelajaran mesin (Zhdanovskaya et al., 2023). Melalui pelabelan, data diberi kategori atau label yang menggambarkan atribut tertentu. Misalnya, dalam analisis kualitas air, atribut Is_safe digunakan untuk menandai keamanan air dengan kelas 0 menunjukkan air tidak aman dan kelas 1 menunjukkan air aman untuk dikonsumsi.

Pembersihan Data

Pembersihan data adalah suatu proses yang digunakan untuk menghapus atau memperbaiki data yang salah, rusak, salah format, duplikat, atau tidak lengkap dari kumpulan data (Cunningham & Muir, 2023; Parulian & Ludäscher, 2023) yang dapat dilihat pada Gambar 2.

```
→ Data Shape : (7996, 21)
-----
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7996 entries, 0 to 7995
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   fluoride   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)
```

Gambar 2. Pemberisihan Data

Gambar 2. menunjukkan hasil proses pembersihan data yang bertujuan untuk mengidentifikasi adanya data yang hilang atau tidak valid. Jika data mengandung nilai NAN (Not A Number) maka data tersebut akan dihapuskan untuk memastikan integritas data sebelum dilakukan pengolahan lebih lanjut. Dalam dataset yang digunakan, teridentifikasi terdapat 4 observasi data yang memuat nilai NAN sehingga data yang digunakan untuk proses lebih lanjut berjumlah 7996 data.

Split Data

Splitting data adalah langkah penting dalam mengembangkan model klasifikasi yang melibatkan pembagian dataset menjadi dua bagian: data latih dan data uji (Rahayu & Yamasari, 2024). Data latih untuk melatih model, sementara data uji untuk mengevaluasi performa model yang dapat dilihat dalam Tabel 2.

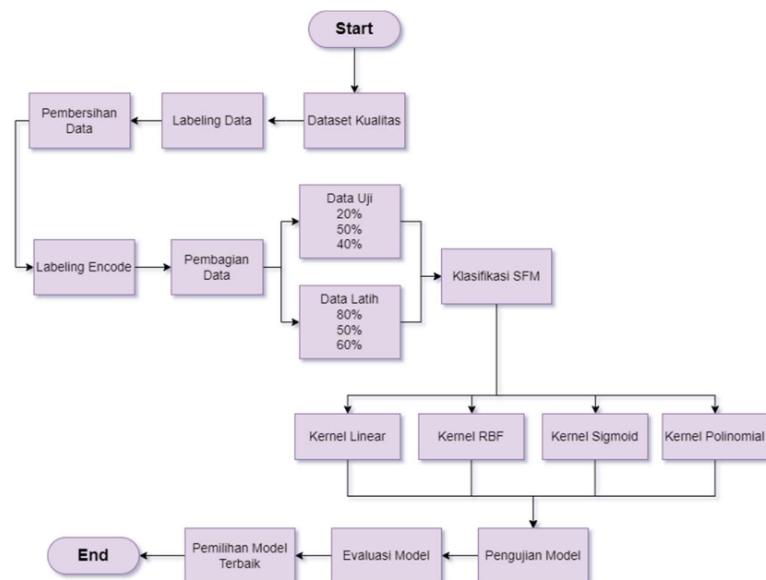
Tabel 2. Pembagian Data

Rasio	Data Latih	Data Uji
80:20	6397	1599
50:50	3889	3998
60:40	4798	3198

Support Vector Machine

Metode SVM sering dipilih sebagai algoritma klasifikasi yang efektif. SVM digunakan untuk memisahkan kelas-kelas dalam data dengan memanfaatkan berbagai jenis kernel seperti linear, polynomial, RBF, dan sigmoid (Nuraeni & Faisal, 2025; Sudin

et al., 2023). Setiap kernel memiliki keunggulan dan kelemahan tertentu dalam menangani pola-pola data yang berbeda. Evaluasi klasifikasi dilakukan dengan mengukur akurasi, presisi, dan recall. Gambar 3 menunjukkan langkah-langkah dalam proses klasifikasi dengan SVM.



Gambar 3. Flowchar Klasifikasi dengan SVM

Algoritma Support Vector Machine (SVM) mendasarkan prinsipnya pada konsep minimisasi risiko struktural (SRM). Fokus utama SVM adalah mencari hyperplane optimal di dalam ruang input yang mampu memisahkan dua kelas data dengan baik (Medyanti & Faisal, 2023). Hyperplane ini ditemukan dengan tujuan mengoptimalkan margin antara kedua kelas, yang diukur dengan mencari titik maksimum dari margin tersebut (Iqbal et al., 2023). SVM memprioritaskan pencarian hyperplane yang memberikan margin terbesar antara kelas-kelas yang ingin dipisahkan, sehingga meminimalkan risiko kesalahan klasifikasi (Nuraeni & Faisal, 2025). Konsep ini memungkinkan SVM untuk menangani dengan efisien permasalahan klasifikasi yang kompleks dan nonlinear dalam berbagai domain aplikasi.

Skenario Pengujian

Dalam penelitian ini, pengujian model dilakukan dengan menggunakan tiga skenario pembagian data yang berbeda untuk mengevaluasi kinerja metode yang digunakan. Pengujian merujuk pada proses evaluasi di mana model yang dikembangkan diuji kemampuannya dalam memprediksi atau mengenali pola pada data baru. Pembagian data yang diterapkan meliputi skenario 80:20, 50:50, dan 60:40. Skenario ini merujuk pada proporsi data yang digunakan untuk pelatihan (training) dan pengujian (testing) model.

Evaluasi Model

Dalam proses evaluasi sistem, terdapat beberapa metrik penting yang digunakan, antara lain akurasi, presisi, dan recall. Presisi dan recall digunakan untuk mengevaluasi seberapa baik sistem dalam mengambil data yang relevan dengan kebutuhan (Sadeli & Lawanda, 2023). Presisi mengukur kecocokan antara elemen data yang diambil dengan informasi yang diperlukan, sementara recall menunjukkan seberapa efektif sistem dalam menemukan semua elemen data yang relevan (Harju & Mesaros, 2023). Berikut ini adalah persamaan yang digunakan dalam menghitung akurasi, presisi dan recall.

$$Akurasi = \frac{TP + FN}{p + n} \quad (1)$$

$$Presisi = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

Hasil dan Pembahasan

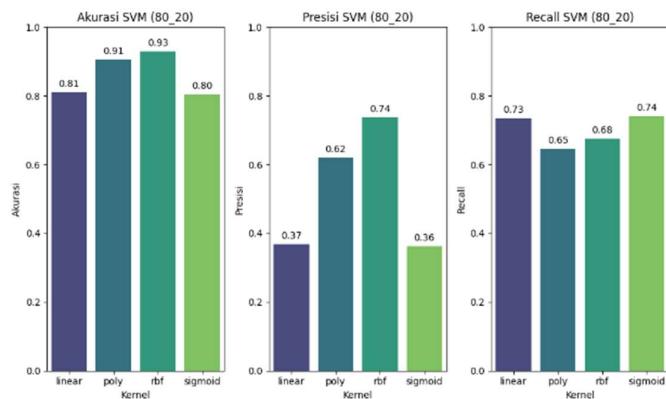
Berdasarkan hasil uji coba dari empat kernel dengan menggunakan 3 sekenario yang telah ditentukan. Hasil dari uji coba akan dibandingkan untuk memahami kinerja relatif dari masing-masing kernel dalam memprediksi kualitas air minum. Hasil evaluasi dapat di lihat pada Tabel 3.

Tabel 3. Hasil Evaluasi

Tabel	Rasio	Akurasi	Presisi	Recall
Linear	80:20	80%	37%	73%
	50:50	79%	34%	77%
	60:40	80	36%	75%
Polynomial	80:20	91%	62%	65%
	50:50	91%	64%	61%
	60:40	91%	65%	65%
RBF	80:20	93%	74%	68%
	50:50	92%	71%	61%
	60:40	93%	74%	64%
Sigmoid	80:20	80%	36%	74%
	50:50	79%	34%	76%
	60:40	80%	35%	75%

Skenario 1

Dalam skenario pertama dengan pembagian data 80:20, hasil evaluasi kinerja model menunjukkan bahwa kernel Linear mencapai tingkat akurasi sebesar 81%, akurasi pada kernel Polynomial mencapai 91%, sedangkan pada kernel RBF mencapai akurasi 93% akurasi, dan kernel sigmoid mencapai tingkat akurasi sebesar 80%.

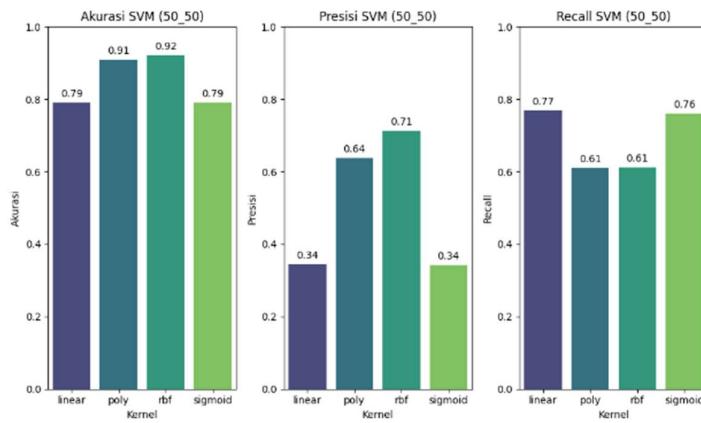


Gambar 4. Evaluasi Skenario 1

Gambar 4 menunjukkan perbandingan performa model klasifikasi menggunakan empat jenis kernel dalam scenario pertama: linear, polynomial, RBF, dan sigmoid. Kernel RBF menunjukkan performa terbaik dengan akurasi 0.93, diikuti oleh kernel polynomial dengan akurasi 0.91. Kedua kernel ini juga memiliki nilai presisi, recall, dan f1-score yang seimbang untuk kedua kelas. Kernel linear dan sigmoid memiliki performa yang lebih rendah, dengan akurasi masing-masing 0.81 dan 0.80. Kernel sigmoid memiliki presisi yang sangat rendah pada kelas 1. Hasil ini menunjukkan bahwa pemilihan kernel mempengaruhi performa model secara signifikan, di mana kernel RBF dan polynomial lebih efektif untuk dataset ini dibandingkan kernel linear dan sigmoid.

Skenario 2

Selanjutnya pada skenario kedua dengan pembagian data 50:50, hasil evaluasi kinerja model menunjukkan bahwa kernel Linear mencapai tingkat akurasi sebesar 79%, akurasi pada kernel Polynomial mencapai 91%, sedangkan pada kernel RBF mencapai akurasi 92%, dan kernel sigmoid mencapai tingkat akurasi sebesar 79%.

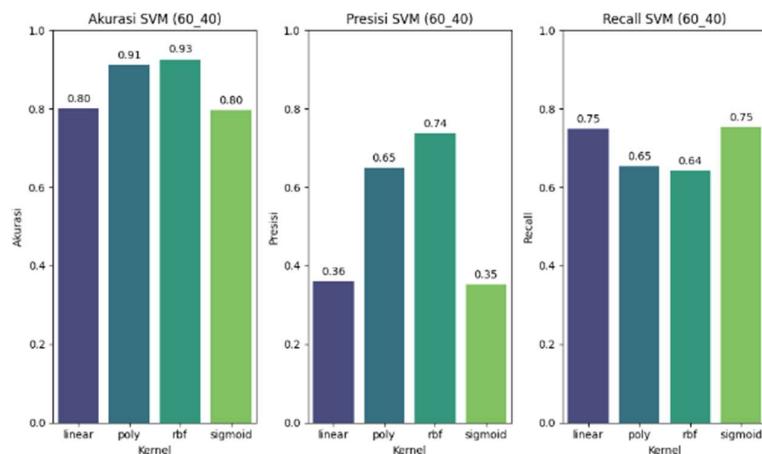


Gambar 5. Evaluasi Skenario 2

Gambar 5 menunjukkan perbandingan performa model klasifikasi dengan empat jenis kernel dalam skenario dua: linear, polynomial, RBF, dan sigmoid. Kernel RBF kembali menunjukkan performa terbaik dengan akurasi 0.92, diikuti oleh kernel polynomial dengan akurasi 0.91. Kedua kernel ini memiliki presisi, recall, dan f1-score yang tinggi dan seimbang untuk kedua kelas. Kernel linear dan sigmoid memiliki akurasi yang lebih rendah, masing-masing 0.79 dan 0.83. Kernel sigmoid kembali menunjukkan presisi yang sangat rendah pada kelas 1. Secara keseluruhan, hasil ini konsisten dengan skenario 1, menunjukkan bahwa kernel RBF dan polynomial lebih unggul dalam menangani dataset ini dibandingkan kernel linear dan sigmoid. Pemilihan kernel yang tepat terbukti sangat mempengaruhi efektivitas model klasifikasi.

Skenario 3

Pada skenario terakhir dengan pembagian data 60:40, hasil evaluasi kinerja model menunjukkan bahwa kernel Linear mencapai tingkat akurasi sebesar 80%, akurasi pada kernel Polynomial mencapai 91%, sedangkan pada kernel RBF mencapai akurasi 93%, dan kernel sigmoid mencapai tingkat akurasi sebesar 80%.



Gambar 6. Evaluasi Skenario 3

Gambar 6. Menunjukkan bahwa kernel RBF menunjukkan performa terbaik dengan akurasi 0.93, memiliki presisi dan recall tinggi untuk kelas 0 (0.95 dan 0.97) serta performa yang cukup baik untuk kelas 1 (0.74 dan 0.64). Kernel polynomial juga efektif dengan akurasi 0.91 dan keseimbangan yang baik antara presisi dan recall untuk kedua kelas. Kernel linear dan sigmoid memiliki akurasi lebih rendah (0.80) dan presisi sangat rendah pada kelas 1 (0.36 dan 0.35). Hasil ini menunjukkan kesamaan pada skenario pertama dan kedua yang menunjukkan bahwa kernel RBF unggul dalam proses analisis data dalam mengklasifikasi kualitas air minum.

Dalam ketiga skenario, evaluasi model klasifikasi menggunakan empat jenis kernel yaitu Linear, Polynomial, RBF, dan Sigmoid menunjukkan bahwa kernel RBF consistently memberikan performa terbaik, dengan akurasi tertinggi mencapai 93% pada skenario pertama dan ketiga, dan 92% pada skenario kedua. Kernel Polynomial juga menunjukkan performa yang sangat baik dengan akurasi stabil di sekitar 91% pada semua skenario. Sebaliknya, kernel Linear dan Sigmoid menunjukkan akurasi yang lebih rendah, dengan masing-masing berkisar di angka 79%-81% dan 79%-80%. Khususnya, kernel Sigmoid

menunjukkan presisi yang sangat rendah pada kelas 1. Secara keseluruhan, hasil ini menegaskan bahwa pemilihan kernel sangat mempengaruhi efektivitas model klasifikasi, dengan kernel RBF dan Polynomial terbukti lebih unggul dalam mengklasifikasi kualitas air minum dibandingkan kernel Linear dan Sigmoid.

Kesimpulan dan Saran

Air merupakan senyawa penting bagi kehidupan yang mendukung reaksi kimia dalam sel dan transportasi nutrisi serta oksigen. Krisis air bersih semakin mengkhawatirkan, hanya 1% air bersih yang bisa dikonsumsi. Penelitian ini berhasil menunjukkan bahwa metode Support Vector Machine (SVM) efektif untuk mengklasifikasikan kualitas air minum. Penggunaan berbagai kernel pada SVM, terutama kernel RBF, menghasilkan akurasi yang tinggi dalam memprediksi kualitas air, dengan akurasi tertinggi mencapai 95%. Temuan ini sejalan dengan studi sebelumnya yang menunjukkan keunggulan SVM dalam menangani data yang kompleks dan tidak seimbang. Penelitian ini menyarankan penggunaan metode SVM dalam aplikasi praktis untuk memastikan kualitas air minum yang aman dan berkualitas. Selain itu, hasil penelitian ini membuka peluang untuk penelitian lanjutan yang dapat mengeksplorasi penggunaan kernel lain atau kombinasi metode untuk meningkatkan akurasi dan efisiensi klasifikasi kualitas air.

Daftar Pustaka

Aldossary, Y., Alhaddad, S., Ebrahim, M., & Zeki, A. M. (2022). Comparing K-Nearest Neighbors, Random Forest and Naïve Bayes Models to Classify Fetal Health Using Resampling Methods. *2022 International Conference on Data Analytics for Business and Industry (ICDABI)*, 250–254. <https://doi.org/10.1109/ICDABI56818.2022.10041489>

Alijanzadeh Maliji, B., Babayemehr, A., Mehrabani, S., & Aghajanpour, F. (2023). Role of the World Health Organization in Management of Gastrointestinal Diseases Caused by Contaminated Water in Children in the Middle East: A Review Article. *Journal of Pediatrics Review*, 11(1), 59–66. <https://doi.org/10.32598/jpr.11.1.285.2>

Cunningham, S. A., & Muir, J. A. (2023). Data Cleaning. In *The Cambridge Handbook of Research Methods and Statistics for the Social and Behavioral Sciences* (pp. 443–467). Cambridge University Press. <https://doi.org/10.1017/9781009010054.022>

Devgade, P., & Patil, M. (2023). Water, sanitation, and hygiene assessment at household level in the community: A narrative review. *Journal of Datta Meghe Institute of Medical Sciences University*, 18(1), 173. https://doi.org/10.4103/jdmimsu.jdmimsu_453_22

Dhaswadikar Usha Sitaram. (2022). Study of the physio-chemical parameters for testing water: A review. *World Journal of Advanced Research and Reviews*, 14(3), 570–575. <https://doi.org/10.30574/wjarr.2022.14.3.0600>

Harju, M., & Mesaros, A. (2023). Evaluating Classification Systems Against Soft Labels with Fuzzy Precision and Recall. *Detection and Classification of Acoustic Scenes and Events (DCASE)*, September, 46–50.

Iqbal, E. A. M., Kusumawati, R., & Santoso, I. B. (2023). Hybrid Model Transfer Learning

ResNet50 and Support Vector Machine for Face Mask Detection. *International Journal of Advances in Data and Information Systems*, 4(2), 125–134. <https://doi.org/10.25008/ijadis.v4i2.1297> <http://repository.uin-malang.ac.id/16448/>

Kashtanjeva, A., Vehapi, I., Kurteshi, K., & Paçarizi, M. (2022). Determining the Physico-Chemical and Microbiological Parameters of the Water Quality in the Batllava Lake, Kosovo. *Journal of Ecological Engineering*, 23(9), 231–240. <https://doi.org/10.12911/22998993/151906>

Koninck, C., Goetz, V., & Stitou, D. (2023). Theoretical study of a membrane filtration process for water treatment using low temperature solar energy. *MATEC Web of Conferences*, 379, 08003. <https://doi.org/10.1051/matecconf/202337908003>

Mardewi, M., Yarkuran, N., Sofyan, S., Aziz, F., & Adriana, A. N. I. (2023). KLASIFIKASI KATEGORI OBAT MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE. *Journal Pharmacy and Application of Computer Sciences*, 1(1), 27–32. <https://doi.org/10.59823/jopacs.v1i1.19>

Marten Sahalatua Tumanger, R., & Hidayat, N. (2019). Komparasi Metode Data Mining Support Vector Machine dengan Naive Bayes untuk Klasifikasi Status Kualitas Air (Vol. 3, Issue 10). <http://j-ptiik.ub.ac.id>

Medyanti, W. A., & Faisal, M. (2023). Early Prediction System for Employee Attrition Company “XYZ” Using Support Vector Machine Algorithm. *CESS (Journal of Computer Engineering, System and Science)*, 8(2), 429. <https://doi.org/10.24114/cess.v8i2.46494>

Mishra, A., & Tushaus, D. W. (2022). Water Scarcity. In *Legal Analytics* (pp. 105–116). Chapman and Hall/CRC. <https://doi.org/10.1201/9781003215998-9>

Nuraeni, N., & Faisal, M. (2025). Classification of Sleep Disorders Using Support Vector Machine. *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*. <https://doi.org/10.22219/kinetik.v10i1.2054>

Parulian, N. N., & Ludäscher, B. (2023). Trust the Process: Analyzing Prospective Provenance for Data Cleaning. *Companion Proceedings of the ACM Web Conference 2023*, 1513–1523. <https://doi.org/10.1145/3543873.3587558>

Rahayu, S., & Yamasari, Y. (2024). Klasifikasi Penyakit Stroke dengan Metode Support Vector Machine (SVM). *Journal of Informatics and Computer Science*, 05.

Sadeli, A. F., & Lawanda, I. I. (2023). Recall, Precision, and F-Measure for Evaluating Information Retrieval System in Electronic Document Management Systems (EDMS). *Khizanah Al-Hikmah : Jurnal Ilmu Perpustakaan, Informasi, Dan Kearsipan*, 11(2), 231–241. <https://doi.org/10.24252/kah.v11i2a8>

Santoso, I. B., Adrianto, Y., Sensusati, A. D., Wulandari, D. P., & Purnama, I. K. E. (2022). Ensemble Convolutional Neural Networks With Support Vector Machine for Epilepsy Classification Based on Multi-Sequence of Magnetic Resonance Images. *IEEE Access*, 10, 32034–32048. <https://doi.org/10.1109/ACCESS.2022.3159923>

Shemer, H., Wald, S., & Semiat, R. (2023). Challenges and Solutions for Global Water Scarcity. *Membranes*, 13(6), 612. <https://doi.org/10.3390/membranes13060612>

Sudin, A., Salmin, M., Fhadli, M., & Mario Mamonto, A. (2023). KLASIFIKASI KELAYAKAN AIR MINUM BAGI TUBUH MANUSIA MENGGUNAKAN METODE SUPPORT VEKTOR MACHINE DENGAN BACKWARD ELIMINATION. In *Jurnal Jaringan dan Teknologi Informasi* (Vol. 3, Issue 1).

Tanega, M. C., Fajardo, A., & Limbago, J. S. (2023). Analysis of Water Quality for Taal Lake Using Machine Learning Classification Algorithm. *2023 20th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, 397–402. <https://doi.org/10.1109/JCSSE58229.2023.10202046>

Tangkelayuk, A. (2022). Klasifikasi Kualitas Air Menggunakan Metode KNN, Naïve Bayes, dan Decision Tree. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 9(2), 1109–1119. <https://doi.org/10.35957/jatisi.v9i2.2048>

Tatarov, P. (2022). Water in Biological and Food Systems. *Agrobiodiversity for Improving Nutrition, Health and Life Quality*, 6(2). <https://doi.org/10.15414/ainhlq.2022.0028>

Zhdanovskaya, A., Baidakova, D., & Ustalov, D. (2023). Data Labeling for Machine Learning Engineers: Project-Based Curriculum and Data-Centric Competitions. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(13), 15886–15893. <https://doi.org/10.1609/aaai.v37i13.26886>

Zhou, P., Li, Z., Snowling, S., Goel, R., & Zhang, Q. (2019). Short-Term Wastewater Influent Prediction Based on Random Forests and Multi-Layer Perceptron. *Journal of Environmental Informatics Letters*. <https://doi.org/10.3808/jeil.201900010>