

Penerapan Metode SVM pada Klasifikasi Kualitas Air

Salma Sheila Maulina Putri¹, Muhammad Arhami^{2*}, Hendrawaty³

^{1,2,3} Jurusan Teknologi Informasi dan Komputer Politeknik Negeri Lhokseumawe
Jln. B.Aceh Medan Km.280 Buketrata 24301 INDONESIA

^{1*} salmasheilamp@gmail.com

² muhammad.arhami@pnl.ac.id

³ hendrawaty@pnl.ac.id

Abstrak— SVM adalah salah satu metode *learning machine* yang bekerja dengan prinsip *Structural Risk Minimization* (SRM) yang bertujuan untuk menemukan *hyperplane* terbaik yang memisahkan dua buah *class* pada *input space*. Dengan melihat konsep metode SVM yang bekerja dengan menemukan fungsi pemisah optimal yang bisa memisahkan dua set data dari dua kelas yang berbeda. Maka, dari konsep tersebut timbul permasalahan sejauh mana penerapan metode SVM mampu menyelesaikan masalah klasifikasi. Dalam penelitian ini klasifikasi yang dilakukan adalah klasifikasi kualitas air yang akan dinilai berdasarkan WQI (*Water Quality Index*). WQI didasarkan pada ambang batas rekomendasi *World Health Organization* (WHO) dengan Sembilan parameter meliputi PH, *Hardness*, *Solids*, *Chloramines*, *Sulfate*, *Conductivity*, *Organic_Carbon*, *Trihalomathanes*, dan *Turbidity*. Untuk itu, penelitian ini melakukan simulasi menggunakan metode SVM dengan pendekatan *One-Versus-One* (OVO) untuk mengevaluasi kemampuan metode tersebut dalam mengklasifikasi kualitas air. Hasil penelitian menunjukkan bahwa nilai performa SVM kernel *linear* dengan parameter optimum $C = 1000$ adalah 60%. Tingkat performa SVM kernel RBF dengan parameter optimum $C = 1000$ dan $\gamma = 4$ adalah 100%. Tingkat performa SVM kernel polinomial dengan parameter optimum $C = 1000$ dan $h = 5$ adalah 98% dan Tingkat performa SVM kernel sigmoid dengan parameter optimum $C = 1000$ adalah 60%. Sehingga performa metode SVM terbaik untuk melakukan analisis klasifikasi data WQI pada Perumda Tirta Pase adalah kernel RBF dengan tingkat akurasi mencapai 100%.

Kata kunci— *Support Vector Machine*, Klasifikasi, Kualitas Air, *Water Quality Index*

Abstract— SVM is a machine learning method that works with the principle of *Structural Risk Minimization* (SRM) which aims to find the best *hyperplane* that separates two classes in the input space. By looking at the concept of the SVM method that works by finding the optimal separator function that can separate two data sets from two different classes. So, from this concept a problem arises to what extent the application of the SVM method is able to solve classification problems. In this study the classification used is the classification of water quality which will be assessed based on the WQI (*Water Quality Index*). WQI is based on the *World Health Organization* (WHO) recommendation threshold with nine parameters including PH, *Hardness*, *Solids*, *Chloramines*, *Sulfate*, *Conductivity*, *Organic_Carbon*, *Trihalomathanes*, and *Turbidity*. For this reason, this study conducted a simulation using the SVM method with the *One-Versus-One* (OVO) approach to evaluate the ability of this method to classify water quality. The results showed that the performance value of the linear kernel SVM with the optimum parameter $C = 1000$ was 60%. The performance level of the RBF kernel SVM with optimum parameters $C = 1000$ and $\gamma = 4$ is 100%. The polynomial kernel SVM performance level with optimum parameters $C = 1000$ and $h = 5$ is 98% and the sigmoid kernel SVM performance level with optimum parameters $C = 1000$ is 60%. So that the best performance of the SVM method for analyzing WQI data classification at Perumda Tirta Pase is the RBF kernel with an accuracy rate of up to 100%.

Keywords— *Support Vector Machine*, Klasifikasi, Kualitas Air, *Water Quality Index*,

I. PENDAHULUAN

Machine Learning adalah cabang ilmu komputer yang berfokus pada pengembangan algoritma dan model komputasional yang dapat belajar dan meningkatkan kinerjanya secara mandiri dari pengalaman dan data yang diberikan. *Machine Learning* memungkinkan komputer untuk "belajar" dari pola dan informasi yang tersembunyi dalam data [1]. Salah satu metode klasifikasi adalah metode *Support Vector Machine* (SVM). SVM merupakan sistem pembelajaran yang menggunakan ruang hipotesis yang berupa fungsi-fungsi *linear* didalam fitur yang memiliki dimensi tinggi dan dilatih dengan menggunakan algoritma pembelajaran berdasarkan teori optimasi [2]. Air dikatakan senyawa kompleks karena mengandung banyak zat-zat dan mineral-mineral didalamnya. Namun, tidak semua kandungan zat dan mineral yang terdapat di dalam air dapat dicerna oleh tubuh manusia. Selain itu, air juga rentan terkontaminasi oleh zat-zat dan bakteri

berbahaya bagi tubuh manusia. Hal tersebut dapat terjadi dikarenakan tercemarnya sumber air ataupun lingkungan yang berada di sekitar sumber air. Dibutuhkan pengawasan dan pengolahan yang ketat sekitar sumber air sehingga dapat dihasilkan kualitas air bersih yang sesuai dengan standar kualitas air bersih serta memenuhi standar air yang layak untuk dikonsumsi oleh manusia [3].

Water Quality Index (WQI) adalah indikator tunggal kualitas air yang dihasilkan dengan memanfaatkan sejumlah karakteristik yang benar-benar mewakili kualitas air. WQI didasarkan pada ambang batas rekomendasi WHO dan Sembilan parameter digunakan untuk menghitung WQI dengan cara tradisional meliputi PH, *Hardness*, *Solids*, *Chloramines*, *Sulfate*, *Conductivity*, *Organic_Carbon*, *Trihalomathanes*, dan *Turbidity* [4].

Tabel 1. Parameter Kualitas Air

Parameter	WHO Limits
Ph	6.5-8.5

Parameter	WHO Limits
Hardness	200 mg/L
Solids	1000 ppm
Chloramines	4 ppm
Sulfate	1000 mg/L
Conductivity	400 pS/cm
Organic Carbon	10 ppm
Trihalomethanes	80 ppm
Turbidity	5 NTU

Penelitian ini akan melakukan pengujian performa metode SVM pada masalah klasifikasi kualitas air dengan menggunakan data WQI. Evaluasi model menggunakan nilai *accuracy* dan *F1-score*. Proses klasifikasi menggunakan metode *Support Vector Machine* sebelumnya pernah dilakukan oleh Imda Innar Ridho, Galih Mahalisa, dengan judul “Analisis Klasifikasi Dataset Indeks Standar Pencemaran Udara (ISPU) Di Masa Pandemi Menggunakan Algoritma *Support Vector Machine* (SVM)” pada tahun 2023 menunjukkan bahwa hasil pengujian tertinggi dengan menggunakan *support vector machine* untuk analisis klasifikasi tingkat kualitas udara berdasarkan dataset ISPU dengan nilai *accuracy* sebesar dengan akurasi 97% [5].

Penelitian lainnya yang dilakukan oleh Indri Monika Parapat, Muhammad Tanzil Furqon, Sutrisno pada tahun 2018 dengan judul “Penerapan Metode *Support Vector Machine* (SVM) Pada Klasifikasi Penyimpangan Tumbuh Kembang Anak” menunjukkan Nilai akurasi terbaik sebesar 63,11%. Dilihat dari hasil pengujian yang telah dilakukan akurasi yang didapatkan tidak terlalu tinggi. Hal ini disebabkan beberapa faktor, yang pertama adalah data yang tidak rata antar satu kelas ke kelas lainnya. Faktor kedua adalah antara data pada satu kelas yang sama memiliki pola data yang berbeda [6].

Penelitian ini diharapkan dapat bermanfaat dalam memberikan pengetahuan bagaimana kemampuan/performa metode SVM pada data WQI.

II. METODOLOGI PENELITIAN

Parameter yang akan digunakan untuk mengukur kualitas air pada Perumda Tirta Pase memiliki total sepuluh variabel. *Output* yang dihasilkan dari pengukuran kualitas air adalah air yang memenuhi syarat dan air yang tidak memenuhi syarat kelayakan untuk diminum. Adapun sepuluh variabel tersebut terdiri dari *PH*, *Hardness*, *Solid*, *Chloramines*, *Sulfate*, *Conductivity*, *Organic_Carbon*, *Trihalomethanes*, *Turbidity*, dan *potability*. Definisi dari setiap parameter kualitas air [7] adalah sebagai berikut:

1. *PH*. *PH* (*Potensial Hidrogen*) adalah konsentrasi ion hydrogen yang ada dalam suatu larutan, dan memungkinkan menjadi pembeda antara status asam dan basa ataupun media netral seperti air. WHO merekomendasikan batas maksimum *PH* yang diperbolehkan adalah 6,5 hingga 8,5.

2. *Hardness* adalah tingkat kemampuan air untuk mengendapkan sabun melalui kehadiran ion-ion logam berharga seperti kalsium dan magnesium.
3. *Solids* adalah jumlah materi anorganik dan organik yang terlarut dalam air.
4. *Chloramines* atau kloramin adalah bahan disinfektan yang digunakan dalam sistem air umum. Kloramin terbentuk saat ammonia dicampur dengan klorin untuk mengolah air minum.
5. *Sulfate* adalah substansi yang dapat menyebabkan rasa air menjadi pahit dengan tingkat yang tinggi, dan dapat menyebabkan efek pencakar jika dikonsumsi oleh individu yang tidak terbiasa.
6. *Conductivity*. Konduktivitas adalah konsentrasi ion bermuatan positif (kation) dan ion bermuatan negatif (anion) dalam air. Konsentrasi ion yang lebih tinggi meningkatkan konduktivitas listrik air.
7. *Organic_carbon* adalah materi yang berasal dari senyawa yang terurai dalam air.
8. *Trihalomethanes*. Trichloromethanes (THMs) adalah senyawa yang terbentuk dalam air minum, terutama sebagai hasil dari penambahan klorin ke materi organik.
9. *Turbidity* adalah tingkat sejauh mana air menjadi keruh.
10. *Potability*. adalah indikator apakah air tersebut aman untuk diminum atau tidak.

2.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data yang diperoleh dari data pengujian air pada Perumda Tirta Pase yang terdiri dari *PH*, *Hardness*, *Solids*, *Chloramines*, *Sulfate*, *Conductivity*, *Organic_Carbon*, *Trihalomethanes*, dan *Turbidity*. Data tersebut diperoleh dari <https://www.kaggle.com/datasets/adityakadiwal/water-potability>

Contoh data klasifikasi kualitas air dapat dilihat pada tabel 2 berikut.

Tabel 2 Contoh Data Kualitas Air

X1	X2	X3	X4	X5	X6	X7	X8	X9	Y
6,60	212,97	23564,57	6,60	325,35	430,33	14,35	33,40	4,45	0
7,14	153,23	11553,03	8,44	329,40	278,48	11,26	81,71	3,79	0
10,01	210,12	9064,03	7,19	363,56	324,64	10,21	92,64	3,84	1
6,33	197,48	23460,82	6,12	303,55	399,56	18,82	62,22	4,03	1
8,89	224,07	22960,18	8,98	274,68	466,24	9,69	59,57	3,61	1

2.2 Algoritma Support Vector Machine

Support Vector Machine (SVM) dikenalkan pertama kali oleh Vapnik tahun 1992 sebagai salah satu metode *learning machine* yang bekerja dengan prinsip *Structural Risk Minimization* (SRM) yang bertujuan untuk menemukan *hyperplane* terbaik yang memisahkan dua buah *class* pada *input space*. Metode ini menggunakan hipotesis berupa fungsi –

fungsi *linier* dalam sebuah ruang fitur yang berdimensi tinggi, dengan mengimplementasikan *learning* bias yang berasal dari teori pembelajaran statistik[6].

Data pada suatu dataset diberikan variabel x_i , sedangkan untuk kelas pada dataset diberikan variabel y_i . Metode SVM membagi dataset menjadi 2 kelas. Kelas pertama yang dipisah oleh hyperline bernilai 1, sedangkan kelas lainnya bernilai -1.

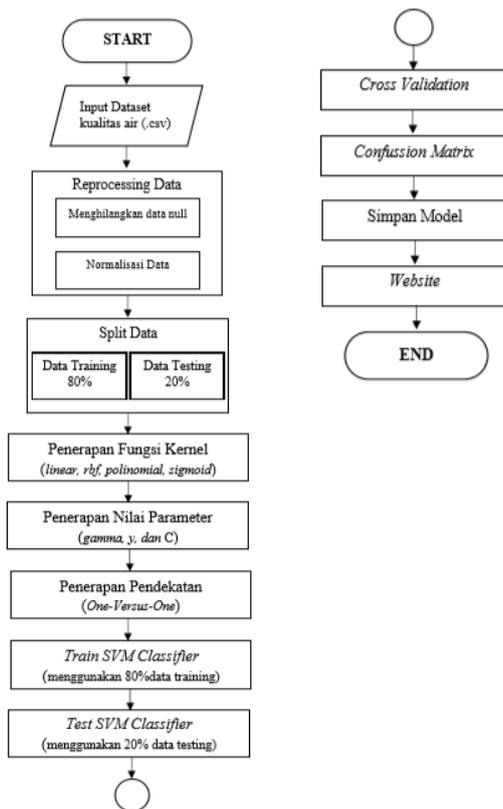
$$x_i \cdot W + b \geq 1 \text{ untuk } y_i = 1 \dots \dots \dots (1)$$

$$x_i \cdot W + b \leq -1 \text{ untuk } y_i = -1 \dots \dots \dots (2)$$

Keterangan:

- x_i = Data ke -i
- w = Nilai bobot *support vector* yang tegak lurus dengan *hyperline*
- b = Nilai bias
- Y_i = Kelas data ke -i

Rancangan implementasi metode SVM pada Klasifikasi Kualitas Air adalah seperti gambar 1 berikut :



Gambar.1 Rancangan Sistem klasifikasi Kualitas Air

Adapun Langkah-langkah tersebut dijelaskan sebagai berikut:

1. *Dataset*. Dataset merupakan data yang akan digunakan sebagai bahan uji penelitian. Pada penelitian ini data yang

diperoleh adalah data uji kualitas air pada perumda Tirta Pase.

2. *Reprocessing Data*. Setelah semua data diperoleh, tidak semua data dapat digunakan. Data yang telah diperoleh perlu melewati tahapan pengolahan awal. Proses utamanya adalah menghilangkan data yang tidak lengkap atau null. Serta normalisasi data.
3. *Split Data*. Tahapan ini akan membagi data menjadi dua bagian yaitu data training dan data testing. Pada penelitian ini data akan dibagi menjadi data training 80% dan data testing 20%.
4. Penerapan fungsi kernel. Fungsi kernel adalah salah satu elemen penting dalam Support Vector Machine (SVM) dalam machine learning. Fungsi kernel digunakan untuk mengubah data ke dalam ruang fitur yang memiliki dimensi yang lebih tinggi, sehingga memungkinkan SVM untuk menemukan batas keputusan yang lebih kompleks dan memisahkan kelas data yang tidak linear terpisah. Kernel yang digunakan pada penelitian ini adalah kernel linear, rbf, polinomial dan sigmoid.
5. Penerapan nilai Parameter. Dalam setiap penerapan kernel terdapat nilai-nilai parameter yang harus ditetapkan, parameter dalam kernel SVM meliputi nilai gamma, y, dan C.
6. *Train SVM Classifier*. Pada tahapan ini model machine learning akan diajarkan pola-pola dalam data. Model diberikan data yang berisi contoh-contoh dengan input dan output yang sesuai. Model menggunakan data tersebut untuk menyesuaikan parameter atau koefisien internalnya sehingga dapat melakukan klasifikasi.
7. *Test SVM Classifier*. Tahapan ini dilakukan untuk menguji performa model Machine Learning setelah proses pelatihan (training) selesai.
8. *Cross Validation*. Dilakukan untuk mengevaluasi performa model prediktif dengan cara membagi dataset menjadi subset yang saling tumpang tindih (folds) untuk pelatihan dan pengujian.
9. *Confussion matrix*. Dilakukan untuk evaluasi kinerja model klasifikasi untuk menggambarkan seberapa baik model tersebut melakukan prediksi terhadap data uji. Matriks ini membantu untuk menghitung metrik evaluasi seperti akurasi, presisi, recall (sensitivitas), spesifisitas, F1-score, dan lainnya.
10. Simpan model. Dilakukan untuk menyimpan parameter dan konfigurasi suatu model setelah melatihnya. Ini memungkinkan untuk menggunakan model SVM yang telah dilatih di ke dalam website yang akan dibuat tanpa harus melalui proses pelatihan lagi.
11. Web Klasifikasi Kualitas Air. Web ini adalah hasil akhir dimana model dengan akurasi terbaik akan digunakan ke dalam web klasifikasi kualitas air.

2.5 Evaluasi Model Menggunakan Confusion Matrix

Confussion matrix adalah salah satu tabel yang berisi perhitungan evaluasi performa model klasifikasi berdasarkan

jumlah objek penelitian yang diprediksi dengan benar dan salah seperti pada tabel 3 berikut.

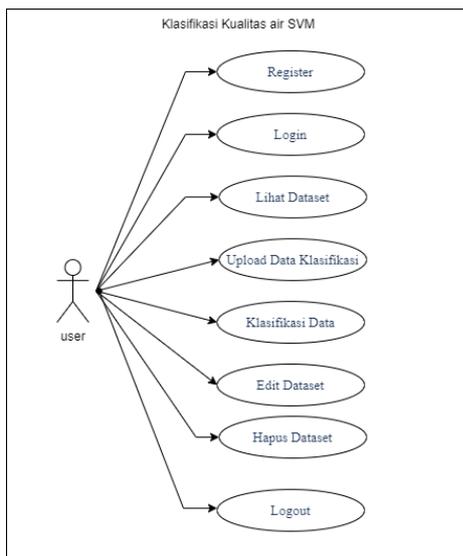
Tabel 3 *Confusion Matrix*

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

Perhitungan *confusion matrix* terdiri dari empat kombinasi nilai prediksi dan nilai aktual. Kombinasi tersebut diistilahkan sebagai *True Positive (TP)*, *True False (TF)*, *False Positive (FP)*, dan *False Negative (FN)*. *True Positive* menunjukkan data positif yang diprediksi benar/tepat. *True Negative* merupakan jumlah dari data negatif yang terdeteksi dengan benar. *False Positive* atau disebut juga *Type-1 Error* merupakan data negatif namun terdeteksi sebagai data yang positif. *False Negative* atau disebut juga *Type-2 Error* menunjukkan data positif tetapi terdeteksi sebagai data *negative*.

2.6 Use Case Diagram

Adapun rancangan *use case* yang digunakan dalam penelitian ini seperti gambar 2 berikut ini berikut :



Gambar 2. Use Case Diagram

Definisi *Use Case* pada sistem klasifikasi kualitas air menggunakan metode SVM dijelaskan dalam table 4 berikut:

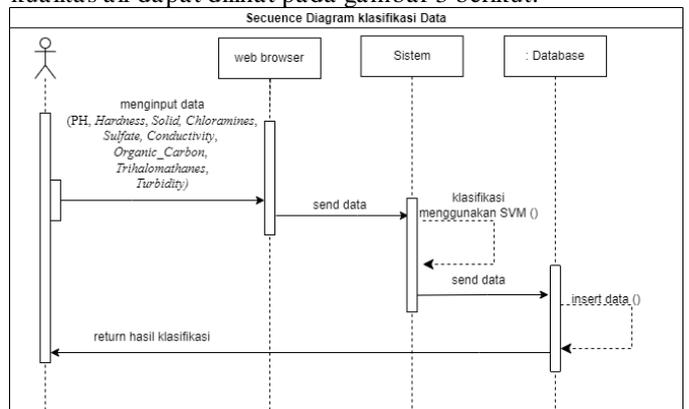
Tabel 4 *Definisi Use Case*

No	Use Case	Deskripsi
1	Register	Proses dimana <i>user</i> dapat melakukan daftar akun pada sistem.
2	Login	Proses dimana <i>user</i> dan admin dapat masuk ke dalam halaman utama sistem.
3	Lihat data Klasifikasi	Proses dimana <i>user</i> dapat melihat data klasifikasi di dalam sistem

No	Use Case	Deskripsi
4	Upload Klasifikasi	Proses dimana <i>user</i> dapat mengupload data klasifikasi baru ke dalam sistem
5	Klasifikasi	Proses dimana <i>user</i> dapat melakukan klasifikasi baru ke dalam sistem
6	Edit dataset	Proses dimana <i>user</i> dapat mengedit data klasifikasi di dalam sistem
7	Delete dataset	Proses dimana <i>user</i> dapat menghapus data klasifikasi di dalam sistem.
8	Logout	Proses dimana admin dan <i>user</i> dapat keluar dari halaman sistem

2.7 Sequence Diagram Klasifikasi Data

Sequence diagram klasifikasi data pada sistem klasifikasi kualitas air dapat dilihat pada gambar 3 berikut.



Gambar 3 *Sequence Diagram* Halaman Klasifikasi Data

Tahapan klasifikasi data yang dijelaskan diatas dimulai dari user menginputkan data *PH, Hardness, Solid, Chloramines, Sulfate, Conductivity, Organic_Carbon, Trihalomathanes, dan Turbidity*. Sistem akan menerima data, kemudian data akan diproses menggunakan metode SVM, disimpan ke dalam data base hingga menampilkan hasil klasifikasi dari kualitas air.

III. HASIL DAN PEMBAHASAN

3.1 Implementasi Support Vector Machine

Hasil klasifikasi Kualitas Air menggunakan metode Support Vector Machine dengan kernel yang digunakan kernel linear, RBF, polynomial, dan sigmoid.

1. Kernel Linear

Hasil performa penggunaan kernel *linear* dari masing-masing nilai C dapat dilihat pada table 5 berikut.

Tabel 5 Performa Klasifikasi SVM Kernel *Linear*

Cost	SVM Kernel <i>Linear</i>			
	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
0.1	60	36	60	45
1	60	36	60	45
10	60	36	60	45
100	60	36	60	45

1000 60 36 60 45

Dapat dilihat bahwa ketika menggunakan metode SVM dengan kernel *linear*, hasil performa nilai *cost* terhadap dataset kualitas air, menunjukkan hal yang sama dalam setiap nilai C dengan nilai akurasi 60%. Selanjutnya adalah penilaian kinerja algoritma *Support Vector Machine*.

	precision	recall	f1-score	support
0	0.60	1.00	0.75	960
1	0.00	0.00	0.00	648
accuracy			0.60	1608
macro avg	0.30	0.50	0.37	1608
weighted avg	0.36	0.60	0.45	1608

Gambar 4 Evaluasi Model Kernel SVM

Hasil evaluasi seperti gambar 4 dapat dilihat bahwa akurasi sebesar 60% dengan nilai precision: negatif 60%, dan positif 00%. Nilai recall: negatif 100%, positif 00%. Nilai f1-score: negatif 75%, positif 00%. Selanjutnya adalah penilaian kinerja algoritma *Support Vector Machine* menggunakan *Confusion Matrix*

Tabel 6 *Confusion Matrix Kernel Linear*

Data Aktual	Data Prediksi	
	Memenuhi Syarat	Tidak Memenuhi Syarat
Memenuhi Syarat (TS)	960	0
Tidak Memenuhi Syarat (TMS)	648	0

2. Kernel RBF

Hasil performa penerapan kernel RBF dengan masing-masing nilai C dan γ dapat dilihat pada table 7 berikut :

Tabel 7 Performa Klasifikasi SVM kernel RBF

SVM Kernel RBF					
Cost	Gamma	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
1	1	99	99	99	99
	2	100	100	100	100
	3	100	100	100	100
	4	100	100	100	100
2	1	100	100	100	100
	2	100	100	100	100
	3	100	100	100	100
	4	100	100	100	100
3	1	100	100	100	100
	2	100	100	100	100
	3	100	100	100	100
	4	100	100	100	100
4	1	100	100	100	100
	2	100	100	100	100
	3	100	100	100	100
	4	100	100	100	100

Tabel 7 menunjukkan bahwa metode SVM menggunakan kernel RBF menghasilkan nilai performa paling yang sangat baik hingga mencapai akurasi 100%. Selanjutnya adalah penilaian kinerja algoritma *Support Vector Machine* menggunakan kernel RBF.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	960
1	1.00	1.00	1.00	648
accuracy			1.00	1608
macro avg	1.00	1.00	1.00	1608
weighted avg	1.00	1.00	1.00	1608

Gambar 3.2 Evaluasi Model Kernel RBF

Hasil evaluasi memperoleh akurasi sebesar 100% dengan nilai precision: negatif 100%, dan positif 100%. Nilai recall: negatif 100%, positif 100%. Nilai f1-score: negatif 100%, positif 100%. Selanjutnya adalah penilaian kinerja algoritma *Support Vector Machine* menggunakan *Confusion matrix*.

Tabel 8 *Confusion Matrix Kernel RBF*

Data Aktual	Data Prediksi	
	Memenuhi Syarat	Tidak Memenuhi Syarat
Memenuhi Syarat (TS)	960	0
Tidak Memenuhi Syarat (TMS)	0	648

3. Kernel Polinomial

Hasil performa penerapan kernel polinomial dengan masing-masing nilai C dan γ dapat dilihat pada Tabel 9 berikut

Tabel 9 Performa Klasifikasi kernel Polinomial

SVM Kernel Polinomial					
Cost	Degree	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
1	1	60	36	60	45
	2	69	72	69	66
	3	66	72	66	58
	4	76	81	76	74
	5	77	82	74	70
10	1	60	36	60	45
	2	70	73	70	66
	3	67	69	67	62
	4	83	86	83	82
	5	85	88	85	85
100	1	60	36	60	45
	2	70	73	70	67
	3	89	90	89	89
	4	88	89	88	88
	5	93	94	93	93
1000	1	60	36	60	45
	2	70	73	70	67
	3	67	69	60	58
	4	90	90	90	90
	5	98	98	98	98

Tabel 9 menunjukkan bahwa metode SVM menggunakan kernel *Polinomial* menghasilkan nilai performa paling baik pada percobaan menggunakan nilai $C = 1000$ dan $h=5$ dengan akurasi mencapai 98%. Selanjutnya adalah penilaian kinerja algoritma *Support Vector Machine*.

	precision	recall	f1-score	support
0	0.96	1.00	0.98	960
1	1.00	0.94	0.97	648
accuracy			0.98	1608
macro avg	0.98	0.97	0.97	1608
weighted avg	0.98	0.98	0.98	1608

Gambar 5 Evaluasi Model Kernel Polinomial

Hasil evaluasi seperti gambar 5 akurasi sebesar 98% dengan nilai *precision*: *negative* 96%, dan *positive* 100%. Nilai *recall*: *negative* 100%, *positive* 94%. Nilai *f1-score*: *negative* 98%, *positive* 97%. Selanjutnya adalah penilaian kinerja algoritma *Support Vector Machine* menggunakan *Confusion matrix*.

Tabel 10 *Confusion Matrix* Kernel Polinomial

Data Aktual	Data Prediksi	
	Memenuhi Syarat	Tidak Memenuhi Syarat
Memenuhi Syarat (TS)	960	0
Tidak Memenuhi Syarat (TMS)	39	609

4. Kernel Sigmoid

Hasil performa penggunaan kernel sigmoid dari masing-masing nilai C dapat dilihat pada tabel berikut.

Tabel 11 Performa Klasifikasi Kernel Sigmoid

SVM Kernel Sigmoid				
Cost	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
0.1	60	36	60	45
1	49	45	49	46
10	45	44	45	44
100	48	48	48	48
1000	47	47	47	47

Tabel diatas menunjukkan bahwa metode SVM menggunakan kernel sigmoid menghasilkan nilai performa terbaik pada percobaan menggunakan nilai $C = 0.1$ dengan akurasi mencapai 60%. Selanjutnya adalah penilaian kinerja algoritma *Support Vector Machine*.

	precision	recall	f1-score	support
0	0.60	1.00	0.75	960
1	0.00	0.00	0.00	648
accuracy			0.60	1608
macro avg	0.30	0.50	0.37	1608
weighted avg	0.36	0.60	0.45	1608

Gambar 6 Evaluasi Model Kernel Sigmoid

Hasil evaluasi seperti gambar 6 akurasi sebesar 60% dengan nilai *precision*: *negative* 60%, dan *positive* 0%. Nilai *recall*: *negative* 100%, *positive* 00%. Nilai *f1-score*: *negative* 75%, *positive* 00%. Selanjutnya adalah penilaian kinerja algoritma *Support Vector Machine* menggunakan *Confusion matrix*.

Tabel 12 *Confusion Matrix* Kernel Sigmoid

Data Aktual	Data Prediksi	
	Memenuhi Syarat	Tidak Memenuhi Syarat
Memenuhi Syarat (TS)	959	1
Tidak Memenuhi Syarat (TMS)	648	0

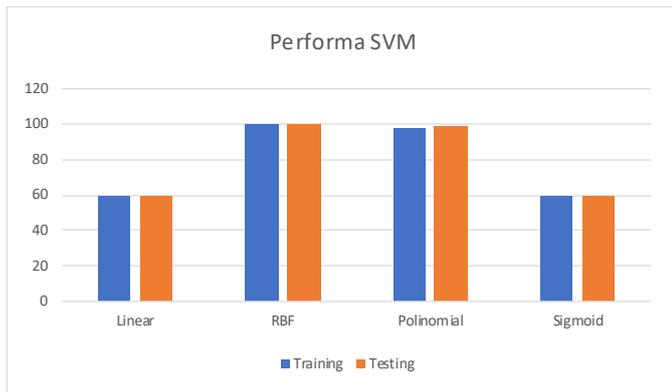
3.2 Perbandingan Performa Klasifikasi SVM

Perbandingan performa klasifikasi dari penerapan metode SVM pada setiap kernel dapat dilihat pada tabel 13 berikut.

Tabel 13 Perbandingan performa klasifikasi SVM

Mode Klasifikasi	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Kernel Linear	60	35	60	44
Kernel RBF	100	100	100	100
Kernel Polinomial	98	98	98	98
Kernel Sigmoid	60	36	60	45

Dalam pengujian yang dilakukan akurasi terbaik adalah penerapan metode SVM dalam penerapan kernel RBF dengan akurasi mencapai 100%, selanjutnya adalah kernel *polinomial* yang memiliki akurasi yang cukup baik pula dengan akurasi mencapai 98%, sedangkan kernel *linear* dan *sigmoid* memiliki akurasi yang kurang baik dengan akurasi sebesar 60%. Korelasi performa pada proses *training* dan *testing* dapat dilihat pada gambar 7 berikut.

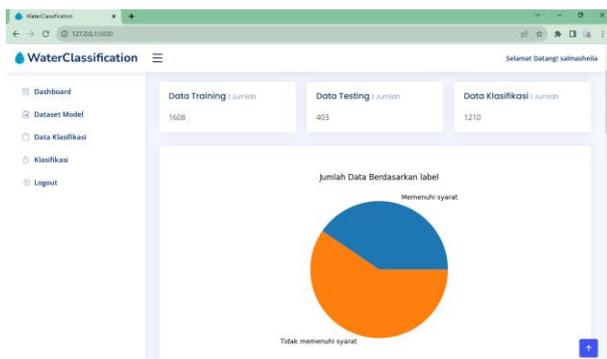


Gambar 7 Perbandingan Performa Klasifikasi SVM

3.3 Implementasi Sistem

1. Halaman Beranda

Halaman beranda adalah halaman utama atau halaman awal yang akan dijumpai oleh *user* setelah berhasil melakukan *login*. Halaman ini berfungsi untuk memberikan gambaran umum tentang apa saja yang dapat dilakukan dalam sistem yang diakses oleh *user*. Implementasi halaman beranda dapat dilihat pada gambar 8 berikut.



Gambar 8 Halaman Beranda

2. Halaman Klasifikasi

Halaman klasifikasi data adalah halaman yang dapat digunakan *user* untuk melakukan klasifikasi kualitas air. Halaman ini merupakan proses utama yang menjadi tujuan dari penelitian ini. Pada proses ini *user* harus menginputkan semua data yang diminta oleh *user* mulai dari *PH*, *Hardness*, *Solid*, *Chloramines*, *Sulfate*, *Conductivity*, *Organic_Carbon*, *Trihalomethanes*, *Turbidity*. *User* dapat memasukkan semua nilai yang menjadi parameter kualitas air, kemudian menekan tombol klasifikasi. Setelah data yang diinputkan oleh *user* diterima dan di proses oleh sistem menggunakan metode *Support Vector Machine*, maka hasil klasifikasi akan tampil pada halaman yang sama.

Halaman klasifikasi pada sistem klasifikasi kualitas air dapat dilihat pada gambar 9 berikut.

Gambar 9 Halaman Klasifikasi Data

IV. KESIMPULAN

Hasil penelitian menunjukkan bahwa nilai performa SVM kernel linear dengan parameter optimum $C = 1000$ adalah 60%. Tingkat performa SVM kernel RBF dengan parameter optimum $C = 1000$ dan $\gamma = 4$ adalah 100%. Tingkat performa SVM kernel polinomial dengan parameter optimum $C = 1000$ dan $h = 5$ adalah 98% dan Tingkat performa SVM kernel sigmoid dengan parameter optimum $C = 1000$ adalah 60%. Sehingga performa metode SVM terbaik untuk melakukan analisis klasifikasi data WQI pada Perumda Tirta Pase adalah kernel RBF dengan tingkat akurasi mencapai 100%.

REFERENSI

- [1] D Kurniawan, *Pengenalan Machine Learning dengan Python*, 1st ed., vol. 1. Elex Media Komputindo., 2002.
- [2] M. A. Rahman, N. Hidayat, and A. Afif Supianto, "Komparasi Metode Data Mining K-Nearest Neighbor Dengan Naïve Bayes Untuk Klasifikasi Kualitas Air Bersih (Studi Kasus PDAM Tirta Kencana Kabupaten Jombang)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* Vol. 2, No. 12, Desember 2018, hlm. 6346-6353 e-ISSN:, vol. 2, no. 12, 2018.
- [3] M. Kencanawati and Mustakim, "Analisis Pengolahan Air Bersih Pada WTP PDAM Prapatan Kota Balikpapan," *Jurnal Transukma*, vol. 02, no. 02, 2017.
- [4] Ade Silvia Handayani, Sopian Soim, Theresia Enim Agusdi, Rumiasih Rumiasih, and Ali Nurdin, "Klasifikasi Kualitas Udara Dengan Metode Support Vector Machine," vol. 3, no. 2, 2020.
- [5] I. I. Ridho and G. Mahalisa, "Analisis Klasifikasi Dataset Indeks Standar Pencemaran Udara (Ispu) Di Masa Pandemi Menggunakan Algoritma Support

- Vector Machine (SVM),” *Technologia : Jurnal Ilmiah*, vol. 14, no. 1, 2023, doi: 10.31602/tji.v14i1.8005.
- [6] I. P. Monika and M. T. Furqon, “Penerapan Metode Support Vector Machine (SVM) Pada Klasifikasi Penyimpangan Tumbuh Kembang Anak,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 10, 2018.
- [7] Md. M. Hassan *et al.*, “Efficient Prediction of Water Quality Index (WQI) Using Machine Learning Algorithms,” *Human-Centric Intelligent Systems*, vol. 1, no. 3–4, 2021, doi: 10.2991/hcis.k.211203.001.