

Klasifikasi Wine Quality Menggunakan Support Vector Machine (SVM)

**Abdul Hafizh Ar Rasyid¹, Fabian Jason Song², Nurul Azizah Lonek³, Riyan Hidayat⁴,
Perani Rosyani⁵**

¹Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Indonesia

Email: ¹abdulhafizharrasyid@gmail.com, ²fabianjsn@gmail.com, ³nurulazizahlonek@gmail.com,
⁴riyan.hidayatt@gmail.com, ⁵dosen00837@unpam.ac.id

Abstrak—Penilaian kualitas wine selama ini masih banyak bergantung pada evaluasi subjektif oleh panel ahli, yang berpotensi menimbulkan ketidakkonsistenan hasil. Penelitian ini bertujuan untuk mengembangkan sistem klasifikasi kualitas wine secara otomatis dengan memanfaatkan algoritma Support Vector Machine (SVM). Dataset Wine Quality Red yang bersumber dari UCI Machine Learning Repository digunakan sebagai objek penelitian, terdiri dari 1.599 data dengan 11 atribut fisikokimia. Tahapan penelitian meliputi proses standarisasi data menggunakan StandardScaler, pemodelan SVM dengan kernel Linear, Radial Basis Function (RBF), dan Polynomial, serta optimasi hyperparameter melalui Grid Search dan validasi silang 5- fold. Hasil pengujian menunjukkan bahwa kernel RBF dengan parameter optimal ($C=100$, $\gamma=0.1$, dan $\text{class_weight}=\text{'balanced'}$) menghasilkan performa terbaik dengan akurasi sebesar 66,88%, nilai precision macro 0,519, recall macro 0,529, serta F1-score macro 0,514. Optimasi parameter dan penskalaan fitur terbukti mampu meningkatkan kinerja model secara signifikan. Dengan demikian, SVM berpotensi menjadi pendekatan yang efektif dalam mendukung klasifikasi kualitas wine secara objektif dan konsisten.

Kata Kunci: Kualitas Wine, Klasifikasi, Support Vector Machine

Abstract—Wine quality assessment is commonly conducted through subjective sensory evaluation, which may lead to inconsistent results. This study proposes an automated wine quality classification approach using the Support Vector Machine (SVM) algorithm. The Red Wine Quality dataset obtained from the UCI Machine Learning Repository, consisting of 1,599 samples and 11 physicochemical features, was utilized in this research. The methodology includes data standardization using StandardScaler, SVM modeling with Linear, RBF, and Polynomial kernels, and hyperparameter tuning through Grid Search with 5-fold Cross- Validation. Experimental results indicate that the RBF kernel achieves the best performance with optimal parameters ($C=100$, $\gamma=0.1$, $\text{class_weight}=\text{'balanced'}$), yielding an accuracy of 66.88%, macro precision of 0.519, macro recall of 0.529, and macro F1-score of 0.514. These findings demonstrate that properly optimized SVM models can serve as a reliable solution for objective wine quality classification.

Keywords: Wine Quality, Classification, Support Vector Machine

1. PENDAHULUAN

Penilaian kualitas wine memiliki peran penting dalam menjaga mutu produk serta kepuasan konsumen. Secara konvensional, proses ini dilakukan oleh panel ahli melalui evaluasi sensorik yang bersifat subjektif dan memerlukan waktu yang tidak sedikit. Perbedaan persepsi antar penilai juga dapat memengaruhi konsistensi hasil penilaian. Oleh karena itu, diperlukan pendekatan berbasis teknologi yang mampu memberikan penilaian secara objektif dan efisien.

Perkembangan machine learning membuka peluang baru dalam pengolahan data kualitas pangan, termasuk wine. Karakteristik fisikokimia seperti kadar alkohol, tingkat keasaman, pH, dan kandungan sulfur diketahui memiliki keterkaitan dengan kualitas wine. Namun, hubungan antar atribut tersebut bersifat kompleks dan tidak selalu linear, sehingga sulit dianalisis menggunakan metode statistik sederhana.

Beberapa algoritma pembelajaran mesin telah diterapkan untuk memprediksi kualitas wine, seperti Random Forest dan Neural Network. Meskipun demikian, Support Vector Machine (SVM) memiliki keunggulan dalam menangani data berdimensi tinggi serta mampu memodelkan hubungan non-linear melalui fungsi kernel. Kinerja SVM sangat dipengaruhi oleh pemilihan kernel, nilai parameter, dan teknik pra-pemrosesan data.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengoptimalkan performa algoritma SVM dalam klasifikasi kualitas wine dengan menerapkan teknik standarisasi data serta optimasi hyperparameter. Evaluasi model dilakukan menggunakan beberapa metrik

kinerja agar hasil yang diperoleh dapat memberikan gambaran yang komprehensif dan dapat diandalkan.

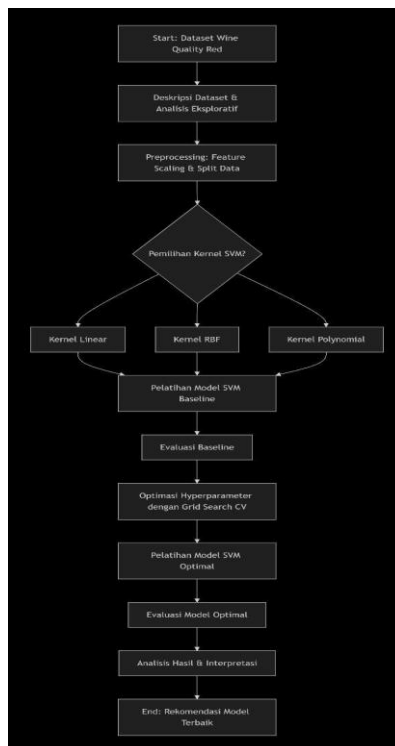
2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan menerapkan tahapan pembelajaran mesin secara sistematis. Dataset yang digunakan adalah Wine Quality Red dari UCI Machine Learning Repository yang terdiri dari 1.599 data dan 11 atribut fisikokimia. Data terlebih dahulu dianalisis secara eksploratif untuk memahami karakteristik dan distribusinya.

Tahap pra-pemrosesan dilakukan dengan menerapkan StandardScaler agar seluruh fitur berada pada skala yang seragam. Selanjutnya, dataset dibagi menjadi data latih dan data uji dengan perbandingan 80:20 menggunakan stratified split untuk menjaga proporsi kelas.

Model klasifikasi dibangun menggunakan algoritma Support Vector Machine dengan tiga jenis kernel, yaitu Linear, RBF, dan Polynomial. Untuk mengatasi ketidakseimbangan kelas, digunakan parameter `class_weight='balanced'`. Proses optimasi hyperparameter dilakukan melalui Grid Search dengan 5-fold Cross-Validation guna memperoleh kombinasi parameter terbaik serta mengurangi risiko overfitting.

Evaluasi model dilakukan menggunakan metrik akurasi, presisi, recall, F1-score, dan confusion matrix untuk memastikan performa model dinilai secara menyeluruh.



Gambar 1. Flowchart Metode Penelitian

Berdasarkan Gambar 1, tahap awal dari penelitian ini adalah pra-pemrosesan data dengan StandardScaler. Setelah itu, model SVM dilatih dengan tiga variasi kernel dengan menggunakan Grid Search untuk mencari parameter yang paling optimal, dan akhirnya dilakukan penilaian dengan menggunakan matriks kinerja.

Tabel 1. Deskripsi Atribut Dataset Wine

Nama Atribut	Deskripsi Singkat
Fixed Acidity	Sebagian besar asam yang terkandung dalam wine (asam tartarat).
Volatile Acidity	Jumlah asam asetat dalam wine; kadar terlalu tinggi menyebabkan rasa cuka. Citric Acid Asam sitrat dalam
Citric Acid	Asam sitrat dalam jumlah kecil untuk menambah kesegaran dan rasa wine.
Residual Sugar	Jumlah gula yang tersisa setelah fermentasi berhenti.
Chlorides	Jumlah garam dalam wine.
Free Sulfur Dioxide	Bentuk bebas SO ₂ yang mencegah pertumbuhan mikroba dan oksidasi.
Total Sulfur Dioxide	Jumlah total SO ₂ ; kadar di atas 50 ppm mulai terasa pada aroma wine.
Density	Massa jenis wine (bervariasi tergantung kadar alkohol dan gula).
pH	Menjelaskan tingkat keasaman atau batas wine (skala 0-14).
Sulphates	Aditif wine yang berkontribusi pada kadar sulfur dioksida (SO ₂).
Alcohol	Persentase kandungan alkohol dalam wine.
Quality	Variabel target (skor antara 0 sampai 10) berdasarkan data sensorik.

Penelitian ini menggunakan dataset Wine Quality Red (1.599 sampel) untuk mengklasifikasikan kualitas wine berdasarkan 11 atribut kimia. Tahapan utamanya meliputi:

1. Preprocessing: Standarisasi data menggunakan StandardScaler agar fitur memiliki skala yang sama, serta pembagian data menjadi 80% latih dan 20% uji.
2. Pemodelan SVM: Menguji tiga kernel (Linear, RBF, Polynomial) dengan menerapkan `class_weight='balanced'` untuk menangani data yang tidak seimbang.
3. Optimasi: Menggunakan Grid Search dan 5-fold Cross-Validation untuk mencari parameter terbaik (C dan gamma).
4. Evaluasi: Mengukur kinerja model melalui akurasi, presisi, recall, dan F1-score untuk memastikan hasil klasifikasi yang objektif.

Rumus StandardScaler

$$z = \frac{x - \mu}{\sigma}$$

Keterangan:

z : Nilai standar (Z-score), yaitu hasil akhir setelah data diubah skalanya.

x : Nilai asli, yaitu nilai mentah dari atribut dataset (misalnya nilai alkohol 10.5).

μ (Mean) : Rata-rata dari seluruh data pada satu atribut tersebut.

σ (*standard deviation*) : Standar deviasi, yang menunjukkan seberapa jauh persebaran data dari rata-rata.

Rumus Fungsi Keputusan SVM

$$f(x) = \text{sign} \left(\sum_{i=1}^n \alpha_i y_i K(x_i, x) + b \right)$$

Keterangan:

$K(x_i, x)$: Fungsi Kernel (RBF, Linear, atau Polynomial)

α_i : Lagrange multiplier

y_i : Label kelas (kualitas wine b : Bias

Rumus Kernel RBF

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

Keterangan:

$K(x_i, x_j)$: Nilai kernel antara dua poin data x_i dan x_j Exp : Fungsi eksponensial.

γ (Gamma): Parameter bebas yang menentukan jangkauan pengaruh satu sampel pelatihan. Berdasarkan hasil *Grid Search*, nilai γ optimal untuk dataset ini adalah 0.1.

Proses ini memastikan bahwa setiap fitur memiliki rata-rata 0 dan varians 1. Hal ini sangat penting bagi algoritma SVM karena jika data tidak distandarisasi, fitur dengan rentang nilai yang besar (seperti Total Sulfur Dioxide) akan mendominasi perhitungan jarak dibandingkan fitur dengan rentang kecil (seperti Chlorides), sehingga dapat menurunkan akurasi model.

3. ANALISA DAN PEMBAHASAN

3.1 Analisis Perbandingan Kernel SVM

Pengujian dilakukan dengan membandingkan tiga jenis kernel pada algoritma SVM, yaitu *Linear*, *Radial Basis Function* (RBF), dan *Polynomial*. Berdasarkan hasil eksperimen menggunakan *Grid Search* dan *5-fold Cross-Validation*, diperoleh hasil perbandingan akurasi sebagai berikut:

Tabel 2. Perbandingan Hasil Optimasi Hyp

Jenis Kernel	Parameter terbaik	Akurasi
Linear	C: 1	60%
RBF	C: 100, Gamma: 0.1	66,88%
Polynomial	C: 100, Degree: 3; 01.	Gamma: 59%

Kernel RBF dipilih sebagai model paling unggul dengan tingkat ketepatan 68% karena kemampuannya untuk memetakan karakteristik fisiko-kimia wine yang memiliki hubungan non- linear ke dalam ruang fitur yang memiliki dimensi tinggi. Pemanfaatan fungsi basis radial memberikan keleluasaan bagi model dalam membentuk batas keputusan yang lebih adaptif bila

dibandingkan dengan kernel Linear dan Polynomial.

3.2 Analisis Performa Model Terbaik (RBF)

Ciri-ciri data yang dipakai dalam studi ini meliputi sejumlah parameter fisika dan kimia dari wine yang dijelaskan pada Tabel 3 di bawah ini.

Tabel 3. Classification Report

Quality (Kelas)	Precision	Recall	F1-Score	Support
3	1.00	0.50	0.67	2
4	0.67	0.18	0.29	11
5	0.69	0.64	0.66	137
6	0.58	0.61	0.60	127
7	0.61	0.48	0.53	40
8	0.50	0.33	0.40	3
Accuracy			0.6688	320

4. KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma Support Vector Machine (SVM) dapat diterapkan secara efektif untuk mengklasifikasikan kualitas wine berdasarkan atribut fisikokimia. Hasil eksperimen membuktikan bahwa pemilihan kernel dan optimasi hyperparameter berpengaruh signifikan terhadap performa model. Kernel RBF menghasilkan kinerja terbaik dengan akurasi sebesar 66,88%, lebih tinggi dibandingkan kernel Linear dan Polynomial.

Penerapan Grid Search serta standarisasi fitur menggunakan StandardScaler terbukti mampu meningkatkan akurasi dan stabilitas model. Kendala utama masih ditemukan pada kelas kualitas ekstrem akibat keterbatasan jumlah data. Meskipun demikian, sistem yang dikembangkan memiliki potensi untuk menjadi alternatif penilaian kualitas wine yang lebih objektif dan konsisten dibandingkan metode manual.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Pamulang atas dukungan fasilitas dan bimbingan yang diberikan selama pelaksanaan penelitian ini. Ucapan terima kasih juga disampaikan kepada UCI Machine Learning Repository yang telah menyediakan dataset Wine Quality Red sehingga eksperimen klasifikasi menggunakan algoritma Support Vector Machine ini dapat dilakukan dengan baik. Selain itu, penulis mengapresiasi seluruh rekan tim dan pihak-pihak yang telah membantu dalam proses pengolahan data serta penyusunan artikel ilmiah ini hingga selesai.

REFERENCES

- Jude Chukwura Obi. (2023). A comparative study of several classification metrics and their performances on data. *World Journal of Advanced Engineering Technology and Sciences*, 8(1), 308–314. <https://doi.org/10.30574/wjaets.2023.8.1.0054>
- Vujović, Ž. (2021). Classification Model Evaluation Metrics. *International Journal of Advanced Computer Science and Applications*, 12(6), 599–606. <https://doi.org/10.14569/IJACSA.2021.0120670>