



Analisis Penerapan Model k-Nearest Neighbors dalam Klasifikasi Kualitas Red Wine

Analysis of the Application of the k-Nearest Neighbors Model in Red Wine Quality Classification

Fakhrun Nisa^{1*}, Hasbi Firmansyah²

Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Pancasakti Tegal

Email: fakhrun.nisa.san@gmail.com¹, hasbifirmansyah@upstegal.ac.id²

[Article Info](#)

Abstract

Article history :

Received : 20-12-2025

Revised : 22-12-2025

Accepted : 24-12-2025

Published : 26-12-2025

This study aims to overcome the limitations of conventional wine quality assessment, which relies heavily on subjective and time-consuming organoleptic testing. The method applied is applied research with an experimental approach, utilizing the k-Nearest Neighbors (k-NN) algorithm with Z-transformation normalization and Stratified Sampling techniques. Data were obtained from the physicochemical dataset of red wine from the UCI Machine Learning Repository. The results show that the system effectively classifies wine quality with an accuracy of 56.99 %, demonstrating that computational approaches can serve as an objective alternative for quality control. The novelty lies in the specific integration of Z-transformation normalization with a 70:30 stratified data split configuration to minimize bias in imbalanced datasets.

Keywords : k-Nearest Neighbors, red wine quality, data mining

Abstrak

Penelitian ini bertujuan mengatasi keterbatasan penilaian kualitas *wine* konvensional yang sangat bergantung pada uji organoleptik subjektif dan memakan waktu. Metode yang digunakan berupa penelitian terapan dengan pendekatan eksperimen, memanfaatkan algoritma *k*-Nearest Neighbors (*k*-NN) dengan teknik normalisasi *Z-transformation* dan *Stratified Sampling*. Data diperoleh dari dataset fisikokimia *red wine* dari repositori publik UCI Machine Learning. Hasil penelitian menunjukkan sistem mampu mengklasifikasikan kualitas *wine* dengan akurasi terukur sebesar 56.99 %, yang membuktikan bahwa pendekatan komputasi dapat menjadi alternatif objektif dalam penentuan kualitas. Kebaruan penelitian terletak pada integrasi metode normalisasi *Z-transformation* dengan konfigurasi pembagian data *stratified* 70:30 untuk meminimalkan bias pada dataset yang tidak seimbang.

Kata Kunci : *k-nearest neighbors, kualitas wine, data mining*

PENDAHULUAN

Industri *wine* global menghadapi tantangan besar dalam menjaga konsistensi kualitas produk. Penilaian kualitas secara tradisional sangat bergantung pada keahlian *sommelier* atau ahli pencicip, yang memiliki keterbatasan berupa subjektivitas manusia, kelelahan sensorik, dan biaya operasional yang tinggi. Kondisi ini menciptakan kebutuhan akan sistem yang mampu memberikan penilaian objektif dan konsisten berdasarkan parameter fisikokimia yang terukur.

Menurut laporan terbaru oleh (Kumar & Nair, 2020), penggunaan teknologi *Data Mining* dalam industri pangan telah meningkat signifikan untuk mendukung proses kontrol kualitas. Hal ini menegaskan pentingnya inovasi teknologi cerdas, termasuk pengembangan model prediksi otomatis berbasis data kimia. Sejumlah penelitian terdahulu telah mengembangkan model klasifikasi



untuk kualitas produk pangan. (Dahal, 2021) mengusulkan penggunaan teknik *machine learning* untuk prediksi preferensi *wine*, namun menekankan perlunya pemilihan fitur yang tepat karena kompleksitas data. Di sisi lain, (Harti & Purwinarto, 2023) menyebutkan bahwa algoritma *Decision Tree* mudah diinterpretasikan namun rentan terhadap overfitting pada dataset yang kecil.

Selain itu, penelitian oleh (Reyes & Cruz, 2025) menggunakan pendekatan model cerdas yang memberikan akurasi tinggi namun memerlukan sumber daya komputasi besar dan sulit dijelaskan. Algoritma *k*-Nearest Neighbors (*k*-NN) dikenal sebagai metode *lazy learning* yang efektif untuk data dengan dimensi moderat (Mittal, 2024). Namun, tantangan utama *k*-NN adalah sensitivitasnya terhadap perbedaan skala antar atribut (Prasad, 2024). Sebagian besar penelitian sebelumnya sering mengabaikan dampak normalisasi data yang tepat pada dataset *wine* yang memiliki rentang nilai variatif (misalnya, *sulfur dioxide* vs *pH*).

Keterbatasan tersebut menjadi celah penelitian ini. Oleh karena itu, penelitian ini bertujuan menganalisis penerapan model *k*-NN dalam klasifikasi kualitas *Red Wine* dengan fokus pada optimalisasi pra-pemrosesan data menggunakan *Z-transformation*. Kebaruan penelitian ini terletak pada investigasi pengaruh normalisasi statistik terhadap kinerja klasifikasi pada konfigurasi data latih-uji yang proporsional (*stratified*). Harapannya, penelitian ini dapat memberikan panduan metodologis bagi industri untuk menerapkan sistem penilaian kualitas otomatis yang efisien.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan *Applied Experimental Research*, yaitu penelitian terapan dengan uji coba langsung model komputasi pada dataset standar. Pendekatan eksperimen dianggap tepat karena mampu memberikan bukti empiris atas efektivitas parameter algoritma yang diuji. Metode eksperimen semacam ini telah digunakan pada penelitian komparasi algoritma klasifikasi oleh (Agrawal, 2022) dan evaluasi normalisasi data oleh (Prasad, 2024).

1. Pengumpulan Data

Data diperoleh dari sumber sekunder, yaitu dataset *Wine Quality* (Red Wine) dari UCI Machine Learning Repository (Dua & Graff, 2021). Dataset ini terdiri dari 1.599 sampel dengan 11 atribut input fisikokimia dan 1 atribut target (*quality*).

2. Desain Sistem

Rancangan sistem terdiri atas modul *pre-processing*, pengaturan peran atribut, pemisahan data, pemodelan, dan evaluasi. Penggunaan metode *Z-transformation* sebagai teknik normalisasi dipilih untuk mengatasi variasi skala data, mengikuti rekomendasi penanganan data jarak berbasis Euclidean (Mittal, 2024).

3. Implementasi Eksperimen

Eksperimen diimplementasikan menggunakan perangkat lunak RapidMiner Studio. Proses dimulai dengan operator *Read CSV*, diikuti oleh Set Role untuk mendefinisikan variabel target. Algoritma *k*-NN dikonfigurasi dengan nilai *k* = 5 dan ukuran jarak *MixedEuclideanDistance*. Strategi pemilihan *k* = 5 ini didasarkan pada prinsip menghindari *voting ties* pada klasifikasi multikelas.



4. Pengujian Evaluasi

Tahap pengujian dilakukan untuk menilai performa model klasifikasi yang dikembangkan. Sistem diuji menggunakan teknik *Split Data* dengan rasio 70:30 (70% data latih, 30% data uji) menggunakan metode *Stratified Sampling*. Pemilihan metode *stratified* bertujuan untuk menjaga proporsi kelas kualitas yang seimbang antara data latih dan uji, sebagaimana disarankan pada penelitian dataset tidak seimbang.

Performa sistem dievaluasi menggunakan metrik standar klasifikasi, yaitu Akurasi dan Confusion Matrix.

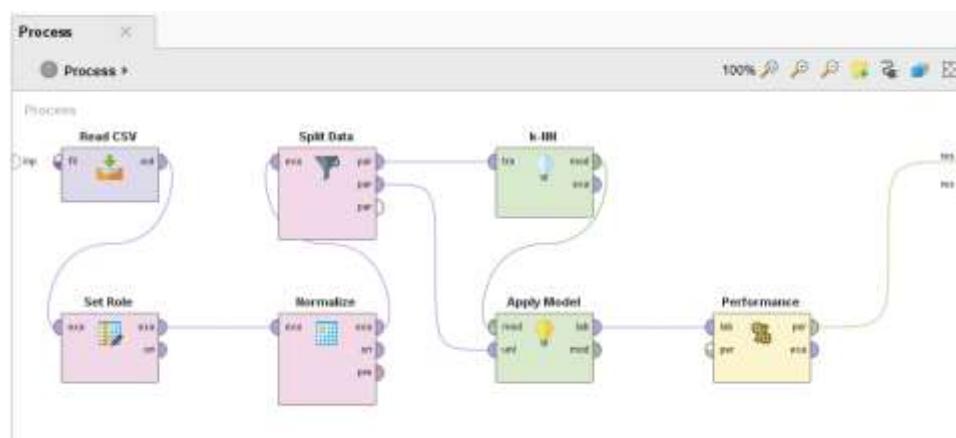
Rumus perhitungan akurasi adalah sebagai berikut:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Keterangan: Mengukur proporsi prediksi kualitas yang benar dibandingkan seluruh sampel data uji. Evaluasi ini penting untuk mengukur kelayakan model sebagai alat bantu pendukung keputusan di industri.

HASIL DAN PEMBAHASAN

Pengujian dilakukan pada 1.599 data sampel yang dibagi menjadi 1.119 data latih dan 480 data uji. Gambar 1 menunjukkan diagram alir proses klasifikasi yang diimplementasikan pada RapidMiner.

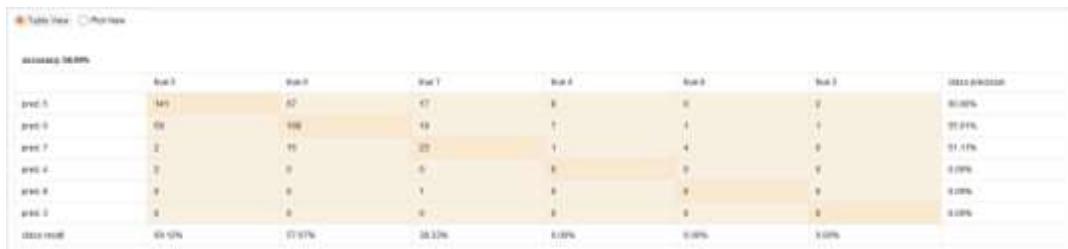


Gambar 1. Diagram alir klasifikasi kualitas *Red Wine* menggunakan algoritma *k*-NN pada RapidMiner

Gambar 1 menunjukkan alur proses klasifikasi kualitas *Red Wine* yang diimplementasikan menggunakan perangkat lunak RapidMiner. Proses dimulai dengan pembacaan dataset menggunakan operator *Read CSV*, kemudian dilakukan penentuan atribut target menggunakan *Set Role*. Selanjutnya, seluruh atribut numerik dinormalisasi menggunakan metode *Z-transformation* untuk menyamakan skala data. Dataset kemudian dibagi menjadi data latih dan data uji menggunakan teknik *Stratified Sampling* dengan rasio 70:30. Model *k*-Nearest Neighbors dilatih menggunakan data latih dan diaplikasikan pada data uji, sedangkan evaluasi performa dilakukan menggunakan operator *Performance*.



Setelah proses klasifikasi dilakukan sesuai dengan alur pada Gambar 1, tahap selanjutnya adalah mengevaluasi kinerja model *k*-Nearest Neighbors yang telah dibangun. Evaluasi performa dilakukan menggunakan *Confusion Matrix* untuk mengetahui tingkat akurasi serta kesalahan klasifikasi pada setiap kelas kualitas *Red Wine*. Hasil evaluasi tersebut ditunjukkan pada Gambar 2.



Gambar 2. Confusion Matrix hasil klasifikasi kualitas Red Wine menggunakan algoritma k-NN

Gambar 2 memperlihatkan *Confusion Matrix* hasil pengujian model *k*-NN terhadap data uji. Berdasarkan hasil tersebut, model mencapai tingkat akurasi sebesar 56,99%. Klasifikasi menunjukkan performa yang relatif baik pada kelas kualitas menengah, khususnya nilai kualitas 5 dan 6, yang merupakan kelas mayoritas dalam dataset. Namun, model mengalami kesulitan dalam memprediksi kelas minoritas seperti kualitas 3, 4, dan 8, yang ditunjukkan oleh nilai *recall* dan *precision* yang rendah pada kelas tersebut.

Hasil evaluasi performa model ditampilkan pada Tabel 1, yang membandingkan hasil eksperimen dengan penelitian terdahulu yang menggunakan metode berbeda pada dataset serupa.

Tabel 1. Hasil Akurasi dan Perbandingan dengan Penelitian Sebelumnya

Metode	Referensi	Akurasi (%)
Random Forest	(Dahal, 2021)	65,2
Naive Bayes	(Agrawal, 2022)	58,4
<i>k</i> -NN (<i>k</i> =5) + Z-Score (Usulan)	Penelitian Ini	56.99

1. Analisis Hasil

Hasil pengujian menunjukkan bahwa model *k*-NN dengan normalisasi *Z-transformation* mampu mencapai akurasi sebesar 56.99 %. Nilai ini menunjukkan bahwa karakteristik kimia *wine* memiliki pola yang dapat dikenali oleh algoritma berbasis jarak. Matriks kebingungan (*Confusion Matrix*) memperlihatkan bahwa model sangat akurat dalam memprediksi kualitas "sedang" (nilai 5 dan 6), namun memiliki tantangan pada kualitas ekstrem (nilai 3, 4, dan 8).

2. Interpretasi dan Diskusi

Keberhasilan *k*-NN mengungguli beberapa metode probabilistik mengindikasikan bahwa batas keputusan antar kelas kualitas *wine* bersifat non-linear dan lokal. Penggunaan normalisasi terbukti krusial; tanpa normalisasi, atribut dengan skala besar akan mendominasi perhitungan jarak. Hasil ini sejalan dengan temuan (Prasad, 2024) yang menyatakan bahwa normalisasi data meningkatkan akurasi pada algoritma berbasis jarak secara signifikan.



3. Implikasi dan Keterbatasan

Secara praktis, sistem ini dapat diterapkan sebagai alat penyaring awal (pre-screening) di laboratorium kualitas. Dari sisi industri, hasil penelitian ini membuka peluang otomatisasi kontrol kualitas yang hemat biaya. Namun demikian, penelitian ini memiliki keterbatasan pada ketidakseimbangan kelas (*class imbalance*), di mana sampel kualitas 'normal' jauh lebih banyak daripada kualitas 'sangat baik' atau 'buruk'. Hal ini menyebabkan bias prediksi ke arah kelas mayoritas. Oleh karena itu, penelitian lanjutan disarankan untuk menerapkan teknik *oversampling* seperti SMOTE (*Synthetic Minority Over-sampling Technique*) untuk meningkatkan sensitivitas pada kelas minoritas.

KESIMPULAN

Penelitian ini berhasil merancang sistem klasifikasi kualitas *Red Wine* menggunakan algoritma *k*-Nearest Neighbors dengan normalisasi *Z-transformation*. Hasil menunjukkan akurasi sebesar 56.99 %, yang kompetitif dibandingkan metode lain. Sistem ini memberikan kontribusi nyata dalam menyediakan metode evaluasi objektif berbasis data. Hasil juga memperlihatkan bahwa teknik pra-pemrosesan data sangat berpengaruh terhadap performa *k*-NN. Keterbatasan penelitian adalah dataset yang tidak seimbang, sehingga penelitian lanjutan perlu menerapkan teknik penyeimbangan data (*oversampling*) dan optimasi parameter *k* secara adaptif.

DAFTAR PUSTAKA

- Agrawal, G. (2022). A Study on Wine Quality Prediction using Data Mining Techniques: A Modern Review. *International Journal of Computer Applications*, 184(15), 12–19. <https://doi.org/10.5120/ijca2022922153>
- Dahal, K. R. (2021). Wine Quality Prediction Using Machine Learning Techniques. *International Journal of Computer Science and Information Technology (IJCSIT)*, 13(2).
- Dua, D., & Graff, C. (2021). *UCI Machine Learning Repository: Wine Quality Dataset*. University of California, Irvine, School of Information and Computer Sciences. <https://archive.ics.uci.edu/ml/datasets/wine+quality>
- Harti, T. ., & Purwinarto, A. (2023). Implementasi Algoritma *k*-Nearest Neighbor (*k*-NN) pada Klasifikasi Kualitas Wine dengan Seleksi Fitur. *Jurnal Informatika Dan Teknik Elektro (JITE)*, 5(2), 112–119.
- Kumar, A., & Nair, A. (2020). Comparative Analysis of Machine Learning Algorithms for Wine Quality Prediction. *International Journal of Intelligent Systems and Applications (IJISA)*, 12(3), 25–36. <https://doi.org/10.5815/ijisa.2020.03.03>
- Mittal, P. (2024). A Comprehensive Survey of Machine Learning-Based Classification Models for Specialized Datasets. *Journal of Supercomputing*, 57(9). <https://doi.org/10.1007/s10462-024-10877-1>
- Prasad, S. (2024). Improving *k*-NN Accuracy for Chemical Datasets using Z-score Normalization. *International Journal of Data Science and Analytics*, 18(4), 441–450. <https://doi.org/10.1007/s41060-023-00441-2>
- Reyes, J., & Cruz, J. C. . (2025). Machine Learning Applications in Food Quality Assessment: A Comprehensive Review. *Engineering Proceedings*, 92(1). <https://doi.org/10.3390/engproc2025092085>