Contents lists available at https://csi.unisi.my.id

Computer Science Insights

Journal Page is available to https://csi.unisi.my.id

Research article

Analisis Kinerja Algoritma K-Nearest Neighbor dan Naive Bayes pada Dataset Wine

Performance Analysis of K-Nearest Neighbor and Naive Bayes Algorithms on the Wine Dataset

*Ayla Zhafira [1], Tri Wahyuni [2], Rahma Yulia Sifa [3]

[1,2,3] Prodi Sistem Informasi, Univeritas Islam Indragiri, Riau, Tembilahan
email: [1] aylanana1144@gmail.com , [2] triwahyuni333@gmail.com, [3] rahmayuliasifa@gmail.com

ARTICLE INFO

Article history:
Received November 11, 2025
Revised November 26, 2025
Accepted November 27, 2025
Available online December 22, 2025

Keywords:
Klasifikasi
K-Nearest Neighbor
*Naive Bayes*Important
Dataset *Wine*
Akurasi

ABSTRACT

Kemajuan dalam teknologi komputasi telah menempatkan klasifikasi sebagai alat esensial dalam pengambilan keputusan otomatis. Studi ini menyajikan analisis komparatif kinerja dua algoritma klasifikasi populer, *K-Nearest Neighbor* (KNN) dan *Naive Bayes*, yang diterapkan pada Dataset *Wine*. Penelitian ini bertujuan menentukan model yang paling optimal dalam hal akurasi dan efisiensi waktu eksekusi. Metode eksperimen kuantitatif diterapkan dengan prapemrosesan StandardScaler serta pembagian data menjadi 70% latih dan 30% uji. Hasil penelitian menunjukkan bahwa *Gaussian Naive Bayes* (GNB) unggul dengan akurasi sempurna 100% dan waktu eksekusi tercepat (0,0037 detik). Sebaliknya, KNN (k = 5) mencapai akurasi 94,44% dengan waktu komputasi yang lebih lambat. Keunggulan GNB ditunjang oleh karakteristik Dataset *Wine* yang sesuai dengan asumsi independensi fitur. Kontribusi penelitian ini terletak pada penyediaan bukti empiris dan panduan praktis mengenai pemilihan algoritma yang efisien pada dataset terstruktur berskala kecil hingga menengah, sehingga dapat menjadi acuan bagi peneliti dan praktisi dalam menentukan model klasifikasi yang tepat pada kasus serupa.

Advancements in computational technology have positioned classification as an essential tool for automated decision-making. This study presents a comparative analysis of the performance of two popular classification algorithms, K-Nearest Neighbor (KNN) and Naive Bayes, applied to the Wine Dataset. The research aims to determine the most optimal model in terms of accuracy and execution efficiency. A quantitative experimental method was employed, incorporating StandardScaler preprocessing and a 70% training–30% testing data split. The results show that Gaussian Naive Bayes (GNB) outperforms KNN with a perfect accuracy of 100% and the fastest execution time (0.0037 seconds). In contrast, KNN (k = 5) achieved an accuracy of 94.44% with slower computational performance. The strong performance of GNB is supported by the characteristics of the Wine Dataset, which align well with the feature independence assumption. The contribution of this study lies in providing empirical evidence and practical guidance for selecting efficient algorithms for small to medium-sized structured datasets, serving as a reference for researchers and practitioners when determining appropriate classification models for similar cases.

1. Introduction

Kemajuan dalam teknologi informasi dan komputasi telah mendorong adopsi teknik pembelajaran mesin secara luas di berbagai bidang analisis data [1]. Dengan kemampuan untuk mengekstraksi pola dari kumpulan data besar dan menghasilkan prediksi yang otomatis, pembelajaran mesin menjadi salah satu solusi penting dalam mendukung proses pengambilan keputusan di bidang industri, kesehatan, bisnis, dan penelitian ilmiah [2] Dalam konteks analisis data, algoritma pembelajaran mesin memungkinkan peneliti untuk mengotomatisasi proses klasifikasi berdasarkan fitur-fitur terukur yang terdapat dalam suatu Dataset. Klasifikasi berperan penting ketika tujuan analisis adalah memprediksi label diskrit pada data baru berdasarkan pola yang dipelajari dari data berlabel, dan kinerja model umumnya diukur menggunakan metrik seperti akurasi, presisi, recall, serta F1-score [3].

Pada penelitian ini, dua algoritma klasifikasi yang populer dan banyak digunakan dalam studi komparatif, yaitu *K-Nearest Neighbor* (KNN) dan *Naive Bayes*, dipilih sebagai objek evaluasi. KNN bekerja dengan menetapkan kelas berdasarkan mayoritas tetangga terdekat dalam ruang fitur, sehingga performanya sangat dipengaruhi oleh

skala dan distribusi data [4]. Sebaliknya, *Naive Bayes* menggunakan pendekatan probabilistik yang mengasumsikan independensi antar fitur, dan dikenal memiliki kecepatan komputasi yang tinggi meskipun karakteristik distribusi data tidak selalu sepenuhnya memenuhi asumsi tersebut [5]. Perbandingan antara kedua algoritma ini telah menjadi topik menarik dalam berbagai penelitian karena perbedaan prinsip kerja dan karakteristik kinerjanya [6].

Dataset *Wine* dipilih sebagai objek eksperimen karena memiliki struktur numerik yang bersih, terdiri atas 178 sampel dengan 13 fitur kimia yang terdefinisi jelas, serta tiga kelas target [7]. Selain itu, Dataset ini tersedia secara siap pakai melalui pustaka scikit-learn dan sering digunakan sebagai data standar dalam penelitian akademik terkait klasifikasi, sehingga memudahkan proses replikasi dan interpretasi hasil [8]. Pada penelitian ini, seluruh proses eksperimen dilakukan menggunakan Google Colab berbasis Python dengan dukungan pustaka numpy, pandas, dan scikit-learn, mulai dari pemuatan data, normalisasi fitur menggunakan StandardScaler, pembagian data menjadi data latih dan data uji (70%–30%), hingga pelatihan dan evaluasi model [9].

Penelitian ini bertujuan membandingkan performa algoritma *K-Nearest Neighbor* (KNN) dan *Gaussian Naive Bayes* dalam melakukan klasifikasi pada Dataset *Wine* melalui evaluasi yang mencakup akurasi, waktu eksekusi, dan analisis confusion matrix untuk memperoleh gambaran kinerja yang komprehensif. Fokusnya adalah menilai kemampuan masing-masing algoritma dalam menghasilkan prediksi yang efektif sekaligus mengidentifikasi metode yang paling optimal berdasarkan kualitas hasil klasifikasi dan kebutuhan waktu komputasi. Dengan pendekatan ini, penelitian diharapkan dapat memberikan wawasan praktis mengenai efektivitas kedua algoritma serta membantu peneliti pemula memahami pertimbangan penting dalam pemilihan metode klasifikasi pada analisis data.

2. Research Methods

2.1 Pendekatan dan Sumber Data

Penelitian ini menggunakan metode eksperimen dengan pendekatan kuantitatif. Fokus penelitian adalah membandingkan performa dua algoritma klasifikasi, yaitu *K-Nearest Neighbors* (KNN) dan *Naive Bayes*, dalam mengklasifikasikan data *Wine* Dataset [10]. Data yang digunakan bersifat sekunder, diperoleh dari pustaka scikit-learn, terdiri atas 178 sampel dengan 13 fitur kimia yang menggambarkan karakteristik anggur. Seluruh proses penelitian dilakukan menggunakan Google Colab berbasis Python dengan dukungan pustaka *numpy*, *pandas*, dan *scikit-learn [11]*.

2.2 Tahapan Penelitian

Tahapan penelitian dilakukan melalui lima langkah utama, yaitu:

a. Pemuatan dan Pemeriksaan Data, yaitu memuat Dataset dan memastikan tidak ada data kosong atau duplikat.
b. Praproses Data, dengan melakukan standarisasi menggunakan *StandardScaler* agar semua fitur memiliki skala yang sama.
c. Pembagian Data, yaitu memisahkan data menjadi 70% untuk pelatihan dan 30% untuk pengujian dengan metode *train_test_split*.
d. Pelatihan Model, yaitu melatih dua algoritma klasifikasi (KNN dan Gaussian *Naive Bayes*) dengan data latih yang telah distandarisasi.
e. Evaluasi Model, yaitu mengukur akurasi, waktu komputasi (fit dan predict), serta menampilkan confusion matrix untuk menilai performa model.

Langkah-langkah tersebut dilakukan secara sistematis untuk memastikan hasil penelitian dapat direplikasi dan dianalisis secara objektif.

2.3 Teknik Analisis Data

Analisis dilakukan dengan pendekatan kuantitatif komparatif untuk menilai performa kedua algoritma berdasarkan hasil pengujian. Nilai akurasi dan metrik lainnya disajikan dalam bentuk tabel dan grafik guna memudahkan interpretasi. Hasil pengujian digunakan untuk menentukan algoritma yang memiliki tingkat efektivitas dan efisiensi terbaik dalam mengklasifikasikan data *Wine* Dataset [12].

3. Results and Discussion

3.1 Hasil Eksperimen

Tahap pengujian dilakukan menggunakan Dataset *Wine* yang terdiri atas 13 fitur numerik serta 3 kelas target yang merepresentasikan jenis anggur berdasarkan karakteristik kimiawinya. Dataset dibagi menjadi dua bagian, yaitu 70% sebagai data latih *(training data)* dan 30% sebagai data uji *(testing data)* menggunakan metode *stratified split* untuk menjaga proporsi kelas yang seimbang. Seluruh proses dilakukan di Google Colab dengan bantuan pustaka *scikit-learn [13]*.

Sebelum pelatihan model, data dinormalisasi menggunakan StandardScaler agar seluruh fitur memiliki rentang nilai yang seragam. Normalisasi ini penting karena algoritma *K-Nearest Neighbor* (KNN) sensitif terhadap

perbedaan skala fitur, sementara *Naive Bayes (NB)* cenderung tetap stabil meskipun tanpa normalisasi [14]. Informasi ringkas mengenai struktur Dataset *Wine* yang digunakan dalam penelitian ini dapat dilihat pada Tabel 1 berikut.

Table 1. Statistik Ringkas Dataset *Wine*

Jumlah Fitur	Jumlah Kelas	Jumlah Data	Data Latih(70%)	Data Uji(30%)
13	3	178	124	54

Setelah proses prapemrosesan selesai, dilakukan pelatihan terhadap dua model klasifikasi, yaitu KNN dengan nilai k = 5 dan Gaussian *Naive Bayes*. Pemilihan nilai k = 5 didasarkan pada praktik umum yang banyak digunakan dalam penelitian serupa, karena jumlah tetangga ganjil dalam skala kecil dapat membantu menjaga keseimbangan antara bias dan varians. Selain mengukur tingkat akurasi, dilakukan juga pencatatan waktu eksekusi (fit dan predict) untuk menilai efisiensi masing-masing algoritma.

3.2 Evaluasi Kinerja Algoritma

Hasil pengujian menunjukkan bahwa kedua algoritma mampu memberikan performa klasifikasi yang sangat baik pada Dataset *Wine*. Nilai akurasi serta waktu eksekusi masing-masing algoritma disajikan secara jelas pada Tabel 2 berikut.

Table 2. Hasil Evaluasi Kinerja Model

Algoritma	Akurasi	Waktu Eksekusi (detik)
K-Nearest Neighbor (k=5)	0.9444	0.0163
Gaussian *Naive Bayes*	1.0000	0.0037

Berdasarkan hasil pada tabel sebelumnya, terlihat bahwa algoritma *Naive Bayes* memberikan kinerja yang sangat optimal dengan akurasi sempurna (100%) serta waktu eksekusi yang jauh lebih singkat, yaitu sekitar 0.0037 detik. Sebaliknya, KNN tetap menunjukkan performa yang tinggi meskipun sedikit lebih lambat, dengan akurasi mencapai 94.44%. Untuk memperoleh gambaran yang lebih detail mengenai pola kesalahan prediksi pada masing-masing algoritma, *confusion matrix* disajikan pada Tabel 3 untuk KNN dan Tabel 4 untuk *Naive Bayes*. Kedua tabel tersebut menjadi dasar dalam melakukan perbandingan menyeluruh terkait kelebihan dan kekurangan masing-masing algoritma.

Table 3. *Confusion Matrix* KNN

	Class 0	Class 1	Class 2
Class 0	18	0	0
Class 1	0	18	3
Class 2	0	0	15

Table 4. *Confusion Matrix Naive Bayes*

	Class 0	Class 1	Class 2
Class 0	18	0	0
Class 1	0	21	0
Class 2	0	0	15

Dari hasil *confusion matrix* terlihat bahwa *Naive Bayes* mampu mengklasifikasikan seluruh data uji dengan benar, tanpa kesalahan prediksi sama sekali. Sementara itu, KNN menunjukkan adanya sedikit kesalahan klasifikasi pada kelas 1, di mana 3 data dari kelas tersebut salah terklasifikasi ke kelas lain.

3.3 Analisis Perbandingan

Berdasarkan hasil yang diperoleh, dapat disimpulkan bahwa kedua algoritma sama-sama menunjukkan performa yang baik dalam pengklasifikasian data *Wine*, namun memiliki karakteristik berbeda dalam hal akurasi dan efisiensi waktu.

Algoritma *Naive Bayes* menunjukkan performa paling tinggi dengan akurasi 100% dan waktu komputasi tercepat. Kemungkinan hasil ini terjadi karena *Naive Bayes* bekerja dengan prinsip independensi antar fitur serta perhitungan probabilitas berbasis distribusi *Gaussian*. Dataset *Wine* yang relatif bersih dan memiliki korelasi antarfitur yang cukup teratur membuat asumsi ini dapat terpenuhi, sehingga hasil prediksinya sangat akurat [15].

Sebaliknya, algoritma KNN memiliki performa yang sedikit lebih rendah, dengan akurasi 94.44%. KNN menentukan kelas berdasarkan kedekatan jarak antar data pada ruang fitur, sehingga performanya sangat dipengaruhi oleh skala fitur dan distribusi data. Walaupun begitu, hasil yang diperoleh tetap menunjukkan bahwa KNN mampu mengenali pola klasifikasi dengan baik [16].

Jika dibandingkan dari sisi efisiensi waktu, *Naive Bayes* juga unggul secara signifikan. Proses komputasinya hanya membutuhkan 0.0037 detik, jauh lebih cepat dibandingkan KNN yang memerlukan 0.0163 detik. Ini menunjukkan bahwa untuk Dataset dengan ukuran kecil hingga menengah, *Naive Bayes* tidak hanya akurat tetapi juga efisien secara waktu.

3.4 Pembahasan Perbandingan

Setelah dilakukan analisis terhadap hasil evaluasi kedua algoritma, tahap selanjutnya adalah melakukan pembahasan perbandingan secara menyeluruh antara *K-Nearest Neighbor* (KNN) dan *Gaussian Naive Bayes* (GNB). Perbandingan ini bertujuan untuk memahami kelebihan, kekurangan, serta karakteristik performa masing-

masing algoritma, baik dari segi akurasi, efisiensi waktu, maupun kemampuan klasifikasi data. Ringkasan perbandingan tersebut disajikan pada Tabel 5, sehingga dapat memberikan gambaran yang lebih komprehensif mengenai algoritma yang paling sesuai diterapkan pada Dataset *Wine*.

Table 5. Perbandingan Kinerja KNN dan *Naive Bayes*

Aspek	*K-Nearest Neighbor* (KNN)	*Gaussian Naive Bayes* (GNB)
Akurasi	94,44%	100%
Waktu Eksekusi	0,0163 detik	0,0037 detik
Kesalahan Klasifikasi	Terdapat 3 kesalahan pada kelas *class_1*	Tidak ditemukan kesalahan klasifikasi
Kelebihan	Adaptif terhadap pola data non-linear dan tidak memerlukan asumsi distribusi data	Cepat, sederhana, serta efisien pada data dengan distribusi normal
Kekurangan	Komputasi lebih berat karena menghitung jarak antar data	Sensitif terhadap pelanggaran asumsi independensi antar fitur

Berdasarkan tabel di atas, dapat disimpulkan bahwa *Naive Bayes* menunjukkan performa yang lebih optimal dibandingkan KNN, baik dari segi akurasi maupun waktu eksekusi. *Naive Bayes* unggul karena memanfaatkan pendekatan probabilistik yang sederhana namun efektif, terutama pada Dataset dengan distribusi normal seperti *Wine* Dataset.

Sementara itu, KNN tetap memberikan hasil yang baik dengan akurasi di atas 90%, meskipun lebih lambat dalam komputasi. Algoritma ini cocok diterapkan pada kasus klasifikasi yang memerlukan fleksibilitas terhadap pola data non-linear. Namun, efisiensinya menurun saat jumlah data meningkat, karena proses perhitungan jarak antar titik menjadi lebih kompleks.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa *Naive Bayes* dapat dianggap sebagai algoritma yang paling efektif dalam klasifikasi Dataset *Wine*, karena mampu memberikan keseimbangan antara kecepatan, kesederhanaan, dan tingkat akurasi yang tinggi.

4. Conclusion

Penelitian komparatif terhadap algoritma *K-Nearest Neighbor* (KNN) dan *Gaussian Naive Bayes* (GNB) pada Wine Dataset menunjukkan bahwa GNB merupakan algoritma paling efektif dengan akurasi 100% dan waktu eksekusi tercepat, sementara KNN memperoleh akurasi 94,44% dengan komputasi yang lebih lambat. Keunggulan GNB dipengaruhi oleh karakteristik Wine Dataset yang bersih dan mendekati distribusi normal sehingga mendukung asumsi independensi fitur. Namun, penelitian ini memiliki keterbatasan karena hanya menggunakan satu dataset dan membandingkan dua algoritma dasar tanpa eksplorasi parameter atau metode validasi yang lebih luas. Oleh karena itu, penelitian selanjutnya disarankan untuk menggunakan berbagai dataset dengan tingkat kompleksitas berbeda, menambah algoritma pembanding, serta menerapkan teknik validasi yang lebih komprehensif agar hasil yang diperoleh lebih general dan robust.

References

[1] A. Rachman Andhika, "Machine Learning Dalam Pengembangan Perangkat Lunak," *Integrative Perspectives of Social and Science Journal*, vol. 2, no. 1, p. 1211, 2025.

[2] N. H. Alhumaidi, D. Dermawan, H. F. Kamaruzaman, and N. Alotaiq, "The Use of Machine Learning for Analyzing Real-World Data in Disease Prediction and Management: Systematic Review," *JMIR Med Inform*, vol. 13, 2025, doi: 10.2196/68898.

[3] P. Nabila, F. Wajidi, and W. Firgiawan, "Analisis Algoritma K-Nearest Neighbor dan Naive Bayes pada Kebijakan Perkuliahan Online dengan Multi Bahasa," *Techno.Com*, vol. 24, no. 2, pp. 524–542, May 2025, doi: 10.62411/tc.v24i2.12656.

[4] S. Arti and E. Suherlan, "Evaluasi Kinerja Machine Learning dalam Memprediksi Kemampuan Adaptasi Mahasiswa pada Lingkungan Pembelajaran Daring," *Jurnal Pustaka AI (Pusat Akses Kajian Teknologi Artificial Intelligence)*, vol. 5, no. 1, pp. 50–57, Apr. 2025, doi: 10.55382/jurnalpustakaai.v5i1.901.

[5] U. Khaira, R. Aryani, and R. W. Hardian, "Komparasi Algoritma Naïve Bayes Dan Support Vector Machine (SVM) Pada Analisis Sentimen Kebijakan Kemdikbudristek Mengenai Kuota Internet Selama Covid-19," *Jurnal PROCESSOR*, vol. 18, no. 2, Nov. 2023, doi: 10.33998/processor.2023.18.2.897.

[6] R. N. Devita, H. W. Herwanto, and A. P. Wibawa, "Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor untuk Klasifikasi Artikel Berbahasa indonesia," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 4, pp. 427–434, Oct. 2018, doi: 10.25126/jtiik.201854773.

[7] Triandes Sinaga, Kevin Bastian Sirait, J. J. Pangaribuan, O. P. Barus, and R. Romindo, "Analisis Kualitas Wine Menggunakan Machine Learning dengan Pendekatan SMOTE dan Seleksi Fitur," *INSOLOGI: Jurnal Sains dan Teknologi*, vol. 4, no. 3, pp. 656–668, Jun. 2025, doi: 10.55123/insologi.v4i3.5436.

[8] E. N. R. Khakim, A. Hermawan, and D. Avianto, "Implementasi Correlation Matrix Pada Klasifikasi Dataset Wine," *JIKO (Jurnal Informatika dan Komputer)*, vol. 7, no. 1, p. 158, Feb. 2023, doi: 10.26798/jiko.v7i1.771.

[9] "Wine Dataset in Sklearn," GeekssforGeeks.

[10] D. Azzahra Nasution, H. H. Khotimah, and N. Chamidah, "Perbandingan Normalisasi Data Untuk Klasifikasi Wine Menggunakan Algoritma K-Nn," *CESS (Journal of Computer Engineering System and*

Science), vol. 4, no. 1, pp. 2502–7131, Jan. 2019, Accessed: Nov. 26, 2025. [Online]. Available: https://www.researchgate.net/

[11] V. Wulandari, W. J. Sari, Z. Alfian, L. Legito, and T. Arifianto, "Implementasi Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor untuk Klasifikasi Penyakit Ginjal Kronik," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 2, pp. 710–718, Apr. 2024, doi: 10.57152/malcom.v4i2.1229.

[12] D. Fabiyanto and Y. Rianto, "Performance Evaluation of Multiple Machine Learning Models for Wine Quality Prediction Evaluasi Kinerja Multiple Model Machine Learning untuk Prediksi Kualitas Wine," *Jurnal Informatika dan Teknologi Informasi*, vol. 21, no. 2, pp. 209–223, 2024, doi: 10.31515/telematika.v21i2.

[13] A. Suarisman, A. Nazir, F. Syafria, and L. Afriyanti, "Perbandingan Jarak Metrik pada Klasifikasi Jamur Beracun Menggunakan Algoritma K-Nearest Neighbor (K-NN)," *Journal of Computer System and Informatics (JoSYC)*, vol. 5, no. 1, pp. 10–19, Nov. 2023, doi: 10.47065/josyc.v5i1.4511.

[14] D. Pratmanto, A. Widayanto, Y. Meisella Kristania, and R. Wijianto, "Analisis Perbandingan Algoritma Naive Bayes Dan KNN Untuk Analisis Sentimen Ulasan Pengguna Aplikasi Vidio Di Google Play Store," *CONTEN : Computer and Network Technology*, vol. 4, no. 2, pp. 119–124, 2024, [Online]. Available: http://jurnal.bsi.ac.id/index.php/conten

[15] R. Eko Putri and R. Rahmawati, "Perbandingan Metode Klasifikasi Naïve Bayes Dan K-Nearest Neighbor Pada Analisis Data Status Kerja Di Kabupaten Demak Tahun 2012," vol. 3, no. 4, pp. 831–838, 2014, [Online]. Available: http://ejournal-s1.undip.ac.id/index.php/gaussian

[16] M. R. Elfansyah, Rudiman, and F. Yulianto, "Perbandingan Metode K-Nearest Neighbor (KKN) dan Naive Bayes Terhadap Analisis Sentimen pada pengguna E-Wallet Aplikasi Dana Menggunakan Fitur Ekstraksi TF-IDF," *Jurnal Keilmuan dan Aplikasi Bidang Teknik Informasi*, vol. 18, no. Vol. 18 No. 2 (2024): Jurnal Teknologi Informasi : Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika, Aug. 2024, doi: https://doi.org/10.47111/jti.v18i2.15009.