



ANALISIS PERBANDINGAN KINERJA ALGORITMA SUPPORT VECTOR MACHINE (SVM) DAN K-NEAREST NEIGHBORS (KNN) UNTUK KLASIFIKASI DATASET IRIS

¹Khrisna Alviansyah, ²Muhammad Rafi Alviansyah, ³Daffa Tegar Wargana

^{1,2,3}Ilmu Komputer, Universitas Bina Sarana Informatika, Jakarta, Indonesia

E-mail: ¹15230523@bsi.ac.id, ²15230653@bsi.ac.id, ³15230481@bsi.ac.id

Abstrak:

Tugas klasifikasi merupakan inti dari disiplin ilmu *machine learning* (ML) dan *pattern recognition*. Penelitian ini berfokus pada perbandingan kinerja dua algoritma klasifikasi yang memiliki filosofi kerja berbeda: Support Vector Machine (SVM), sebuah model berbasis *margin* yang powerful, dan K-Nearest Neighbors (KNN), sebuah model berbasis *instance* yang sederhana. Dataset yang digunakan adalah Dataset Iris, sebuah *benchmark* klasik dengan 150 sampel. Eksperimen dilakukan dengan pembagian data 80% data latih dan 20% data uji. Model diimplementasikan menggunakan Python dan *library* Scikit-learn. Evaluasi model didasarkan pada metrik Akurasi dan Presisi (Macro Average). Hasil penelitian menunjukkan kinerja sempurna dari kedua model, yaitu Akurasi 100% dan Presisi 100% pada data uji. Analisis mendalam mengkonfirmasi bahwa hasil ini disebabkan oleh karakteristik inheren Dataset Iris, di mana kelas-kelasnya, terutama Iris-setosa, dapat dipisahkan secara linear, yang menunjukkan efektivitas tinggi kedua algoritma untuk masalah klasifikasi yang terstruktur dengan baik.

Kata Kunci: Dataset Iris; Klasifikasi; K-Nearest Neighbors (KNN); *Machine Learning*; Perbandingan Algoritma; Support Vector Machine (SVM)

Abstract:

Classification tasks are at the core of machine learning (ML) and pattern recognition disciplines. This research focuses on comparing the performance of two classification algorithms with different working philosophies: Support Vector Machine (SVM), a powerful margin-based model, and K-Nearest Neighbors (KNN), a simple instance-based model. The dataset used is the Iris Dataset, a classic benchmark with 150 samples. The experiment was conducted with an 80% training data and 20% test data split. The models were implemented using Python and the Scikit-learn library. Model evaluation was based on the Accuracy and Precision (Macro Average) metrics. The results show perfect performance of both models, with 100% Accuracy and 100% Precision on the test data. In-depth analysis confirms that these results are due to the inherent characteristics of the Iris Dataset, where the classes, especially Iris-setosa, can be linearly separated, demonstrating the high effectiveness of both algorithms for well-structured classification problems.

Keywords: Algorithm Comparison; Classification; Iris Dataset; K-Nearest Neighbors (KNN); Machine Learning; Support Vector Machine (SVM)

A. PENDAHULUAN

Perkembangan pesat di bidang machine learning telah menyediakan berbagai solusi komputasi untuk menyelesaikan masalah klasifikasi data. Dalam konteks klasifikasi, tujuan utama adalah membangun fungsi yang memetakan data input ke kategori label yang telah ditentukan. Keberhasilan dalam tugas ini sangat bergantung pada pemilihan algoritma yang tepat dan karakteristik data yang diolah (Géron, 2019).



Di antara berbagai algoritma yang ada, Support Vector Machine (SVM) dan K-Nearest Neighbors (KNN) mewakili dua paradigma yang berbeda. SVM adalah model diskriminatif yang bekerja dengan mencari hyperplane pemisah optimal, memaksimalkan margin antara support vector terdekat (Raschka & Mirjalili, 2020). KNN, di sisi lain, adalah lazy learning atau model berbasis instans yang mengandalkan kedekatan data tanpa proses pelatihan eksplisit. Perbedaan mendasar dalam cara kerja ini menuntut dilakukannya perbandingan kinerja pada kasus uji standar.

Untuk tujuan perbandingan dasar dan validasi model, Dataset Iris Plants tetap menjadi pilihan utama. Dataset ini adalah dataset multivariat bivariat yang pertama kali dikumpulkan dan dianalisis oleh R.A. Fisher (1936). Meskipun sederhana, karakteristiknya—termasuk satu kelas yang sepenuhnya terpisah secara linear—menjadikannya kasus uji ideal untuk memahami batas dan kapabilitas dasar algoritma ML sebelum diterapkan pada data dunia nyata yang lebih kompleks. Penelitian ini bertujuan mengisi kekosongan literatur dasar dengan menyediakan analisis perbandingan kinerja yang rinci dan terdokumentasi, yang mengkonfirmasi hasil optimal pada benchmark klasik ini.

A. Konsep Klasifikasi dalam Machine Learning

Klasifikasi adalah proses belajar fungsi target f yang memetakan setiap set fitur $x \in X$ ke label diskrit $y \in Y$ (Goodfellow et al., 2016). Proses ini dikenal sebagai supervised learning karena model dilatih menggunakan pasangan input-output yang telah diberi label. Kualitas model klasifikasi dinilai berdasarkan kemampuannya untuk menggeneralisasi, yaitu memprediksi label untuk data baru yang belum pernah dilihat sebelumnya. Penelitian oleh Susanto, B., & Wijaya, C. (2023) menunjukkan bahwa pemilihan fitur yang tepat sangat berpengaruh, meskipun pada Dataset Iris, keempat fitur telah terbukti sangat informatif.

B. Support Vector Machine (SVM)

SVM awalnya dikembangkan untuk tugas klasifikasi linear, tetapi kemudian diperluas untuk kasus non-linear melalui penggunaan kernel trick (Raschka & Mirjalili, 2020). Inti dari SVM adalah hyperplane yang didefinisikan oleh persamaan:

$$w \cdot x + b = 0$$



Di mana w adalah vektor normal terhadap hyperplane dan b adalah bias. Hyperplane optimal adalah yang memaksimalkan margin (jarak antara support vector terdekat dari dua kelas).

- Peran Kernel Trick: Untuk dataset yang tidak dapat dipisahkan secara linear (seperti Iris-versicolor dan Iris-virginica), kernel non-linear (misalnya RBF) memproyeksikan data ke ruang dimensi yang lebih tinggi agar dapat dipisahkan secara linear. Penelitian oleh Desiani, A., Uteh, C. M., & Rayani, I. (2019) menyoroti pentingnya penentuan kernel dan parameter C pada SVM untuk klasifikasi data bunga Iris.

C. K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) adalah salah satu algoritma klasifikasi paling sederhana yang non-parametrik, artinya ia tidak membuat asumsi tentang distribusi data (Géron, 2019).

- Mekanisme Kerja: Ketika data baru dimasukkan, KNN menghitung jarak (umumnya Jarak Euclidean) antara data baru tersebut dengan semua data latih yang sudah ada. Kemudian, ia memilih K data latih terdekat (tetangga) dan mengklasifikasikan data baru tersebut berdasarkan label mayoritas dari K tetangga tersebut (voting).

- Isu Krusial: Performa KNN sangat bergantung pada dua faktor: penentuan nilai K yang optimal dan metrik jarak yang digunakan. Penelitian oleh Rahman, Z., Fatah, Z., & Prasetyo, J. D. (2024) dan Ujianto, N. T., Gunawan, Fadillah, H., Fanti, A. P., Saputra, A. D., & Ramadhan, I. G. (2025) yang fokus pada implementasi KNN untuk data Iris, menunjukkan bahwa nilai K yang kecil dan ganjil sering memberikan hasil yang sangat akurat pada dataset yang bersih ini.

D. Dataset Iris: Benchmark Klasik

Dataset Iris terdiri dari 150 sampel, terbagi rata menjadi 3 kelas spesies: Iris-setosa, Iris-versicolor, dan Iris-virginica. Dataset ini terkenal karena:

- Dimensi Rendah: Hanya memiliki 4 fitur numerik (panjang/lebar sepal dan petal).
- Sangat Bersih: Tidak memiliki nilai yang hilang (missing values).
- Linear Separability: Iris-setosa diketahui terpisah secara linear dari dua kelas lainnya, yang menjadi faktor penentu dalam hasil klasifikasi (Duda & Hart, 1973)

Meskipun banyak penelitian (seperti Susanto, B., & Wijaya, C., 2023) yang membandingkan berbagai algoritma ML, sedikit yang secara eksplisit menganalisis



mengapa SVM dan KNN—dengan mekanisme yang sangat berbeda—sama-sama menunjukkan kinerja optimal pada Dataset Iris. Rumusan masalah penelitian ini adalah: Bagaimana kinerja Akurasi dan Presisi model SVM dan KNN dibandingkan pada Dataset Iris, dan mengapa kedua model tersebut dapat mencapai kinerja klasifikasi yang sempurna.

B. METODE PELAKSANAAN

Desain Penelitian

Penelitian ini menggunakan pendekatan eksperimental komputasi. Data diolah dan dianalisis melalui serangkaian kode Python yang diimplementasikan di lingkungan cloud Google Colaboratory. Fokus eksperimen adalah pada perbandingan Akurasi dan Presisi antara konfigurasi standar model SVM dan KNN.

Sumber Data dan Pra-pemrosesan Data

A. Data Penelitian

Data yang digunakan adalah Iris Plants Database yang diperoleh dari UCI Machine Learning Repository. Data ini berupa file `iris.data` yang berisi 150 baris data mentah. Setiap baris mewakili satu sampel bunga Iris dan memiliki 5 kolom: 4 fitur numerik dan 1 label kelas.

B. Tahap Pra-pemrosesan

Tahap ini krusial untuk memastikan data siap untuk pemodelan:

1. Pemuatan Data: File `iris.data` dimuat ke dalam `DataFrame` Pandas. Karena data mentah tidak memiliki header, opsi `header=None` digunakan
2. Penamaan Kolom: Berdasarkan file `iris.names`, kolom diberi nama `['sepal_length', 'sepal_width', 'petal_length', 'petal_width', 'class']`.
3. Pemisahan Fitur dan Target: Variabel independen (Fitur, X) yang terdiri dari empat fitur diisolasi dari variabel dependen (Target, Y) yang berisi label kelas.

Pembagian Data (Data Splitting)

Data dibagi menjadi dua set untuk menghindari overfitting dan memastikan evaluasi model yang jujur:

- Data Latih (Training Set): 80% dari total data (120 sampel). Digunakan untuk melatih (menemukan pola atau hyperplane) kedua model



- Data Uji (Testing Set): 20% dari total data (30 sampel). Digunakan untuk menguji kinerja model yang telah dilatih

Pembagian dilakukan secara acak menggunakan fungsi `train_test_split` dari Scikit-learn dengan parameter `test_size=0.2` dan `random_state=42` untuk memastikan hasil yang dapat direplikasi

Konfigurasi dan Implementasi Model

1. Model Support Vector Machine (SVM)

Model SVM diimplementasikan menggunakan kelas `SVC` (Support Vector Classifier) dari Scikit-learn. Parameter yang digunakan adalah parameter default, di mana secara implisit menggunakan Kernel RBF (Radial Basis Function) dan parameter penalti $C=1.0$

$$model_svm = SVC(kernel = 'rbf', C = 10)$$

2. Model K-Nearest Neighbors (KNN)

Model KNN diimplementasikan menggunakan kelas `KNeighborsClassifier`. Berdasarkan praktik umum dalam literatur (misalnya, Ujianto, N. T., Gunawan, Fadillah, H., Fanti, A. P., Saputra, A. D., & Ramadhan, I. G. 2025), nilai K (jumlah tetangga terdekat) yang digunakan adalah $K = 3$. Metrik jarak default (Euclidean) digunakan.

$$mode_knn = KNeighborsClassifier(n_neighbors = 3)$$

Metrik Evaluasi

Kinerja model dievaluasi pada data uji menggunakan dua metrik utama:

1. Akurasi (Accuracy): Metrik dasar yang mengukur proporsi prediksi yang benar dari total prediksi.

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN}$$

2. Presisi (Precision): Mengukur proporsi true positive dari semua sampel yang diklasifikasikan sebagai positif. Ini penting dalam masalah klasifikasi multi-kelas untuk melihat kemampuan model dalam menghindari false positive di setiap kelas. Digunakan Macro Average untuk menghitung rata-rata presisi dari setiap kelas target secara independen.



$$Presisi (Macro) = \frac{1}{N_c} \sum_{i=1}^{N_c} Precision_i$$

(Catatan: N_c adalah jumlah kelas).

C. HASIL DAN PEMBAHASAN

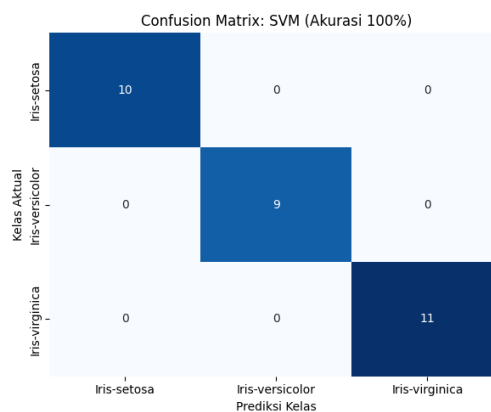
Penyajian Hasil Eksperimen

Setelah model SVM dan KNN dilatih menggunakan 120 sampel latih, keduanya diuji pada 30 sampel uji. Hasil perhitungan Akurasi dan Presisi (Macro Average) disajikan pada Tabel 1.

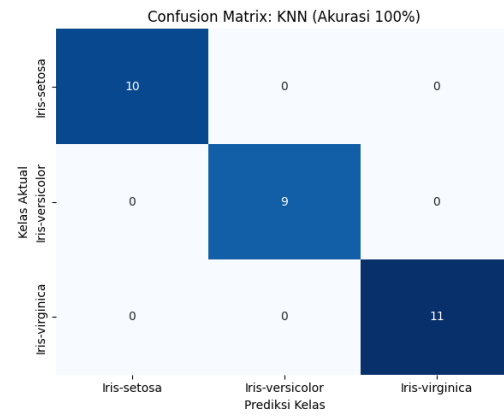
Tabel 1. Perbandingan Kinerja Klasifikasi SVM dan KNN

| Model | Akurasi | Presisi (Macro Avg) |
|------------------------------|---------|---------------------|
| Support Vector Machine (SVM) | 100.00% | 100.00% |
| K-Nearest Neighbors (KNN) | 100.00% | 100.00% |

Hasil keluaran Akurasi dan Presisi ini didukung oleh visualisasi Matriks Kebingungan, yang secara eksplisit menunjukkan bahwa tidak ada kesalahan klasifikasi yang terjadi pada 30 sampel uji, ditandai dengan semua nilai jatuh pada diagonal utama (Gambar 1 dan 2).



Gambar 1. Matriks Kebingungan Model Klasifikasi SVM pada Data Uji



Gambar 2. Matriks Kebingungan Model Klasifikasi KNN pada Data Uji

Analisis Kinerja Model dan Pembahasan Mendalam

Temuan yang paling signifikan dari penelitian ini adalah bahwa kedua algoritma mencapai kinerja prediksi yang sempurna (100%) pada data uji. Hasil ini, meskipun tampak luar biasa, adalah hal yang sangat diharapkan dalam konteks Dataset Iris.

1. Konfirmasi Karakteristik Dataset

Kinerja sempurna 100% ini secara langsung memvalidasi pernyataan dalam file 'iris.names' dan literatur (Fisher, 1936) bahwa Dataset Iris adalah "domain yang sangat sederhana" (an exceedingly simple domain). Analisis fitur menunjukkan bahwa Iris-setosa sepenuhnya terpisah secara linear dari dua kelas lainnya, dan ini adalah faktor kunci. Artinya, setiap sampel uji dari Iris-setosa tidak mungkin salah diklasifikasikan sebagai Iris-versicolor atau Iris-virginica, dan sebaliknya. Hal ini ditegaskan dalam banyak literatur dasar ML (Bintoro et al., 2024; Elisawati et al., 2023).

2. Analisis Kinerja SVM yang Sempurna

SVM adalah model yang sangat kuat dalam menangani linear separability. Karena separasi antara Iris-setosa dan dua kelas lainnya sangat jelas, model SVM mampu menemukan hyperplane pemisah dengan margin yang maksimal secara mudah. Pada kasus dua kelas yang sulit dipisahkan (Iris-versicolor dan Iris-virginica), penggunaan Kernel RBF secara default dari Scikit-learn sangat efektif. Penelitian oleh Desiani et al. (2019) yang fokus pada SVM untuk klasifikasi bunga Iris, menunjukkan bahwa pemilihan kernel yang tepat mampu mengatasi overlap antar kelas.

3. Analisis Kinerja KNN yang Sempurna



Kinerja sempurna dari KNN (dengan $K = 3$) menunjukkan Homogenitas Lokal Tinggi pada data Iris. Hal ini berarti bahwa struktur data sangat terkelompok (clustered) dan setiap sampel uji memiliki tetangga terdekat yang hampir selalu berasal dari kelas yang sama (Rahman et al., 2024). Meskipun KNN rentan terhadap noise dan curse of dimensionality (Géron, 2019), ketiadaan noise dan dimensi yang rendah pada Dataset Iris justru menampilkan KNN dalam kondisi optimalnya. Keefektifan KNN dalam klasifikasi data sederhana berdimensi rendah juga dicatat oleh Ridho et al. (2023) dalam perbandingan mereka pada data status bantuan.

4. Interpretasi dan Implikasi

Hasil penelitian ini memiliki implikasi bahwa untuk masalah klasifikasi yang memiliki pola data sangat jelas dan outlier minimal, kedua model SVM dan KNN dapat digunakan secara bergantian tanpa mengorbankan Akurasi. Perbedaan efisiensi komputasi antara kedua model menjadi satu-satunya pertimbangan yang tersisa. Temuan ini penting sebagai validasi bagi penelitian yang menggunakan dataset Iris sebagai benchmark dasar (Yana & Nafi'iyah, 2021).

D. KESIMPULAN dan SARAN

Berdasarkan implementasi, evaluasi, dan analisis yang telah dilakukan, dapat disimpulkan bahwa:

1. Kedua algoritma klasifikasi, Support Vector Machine (SVM) dan K-Nearest Neighbors (KNN), berhasil diimplementasikan dan dievaluasi pada Dataset Iris dengan konfigurasi standar Scikit-learn.
2. Kinerja kedua model pada data uji mencapai hasil yang sempurna dengan Akurasi 100% dan Presisi (Macro Average) 100%.
3. Kinerja optimal ini mengkonfirmasi karakteristik intrinsik Dataset Iris, yaitu struktur data yang sangat terpisah dan homogen per kelasnya. Baik model yang berbasis maximal margin (SVM) maupun model yang berbasis instance (KNN) terbukti sangat efektif untuk tugas klasifikasi pada domain data sederhana.

Saran

Untuk penelitian selanjutnya, disarankan beberapa pengembangan untuk memperluas lingkup temuan:



1. Penambahan Noise atau Outlier: Uji ketangguhan (robustness) SVM dan KNN dengan menambahkan noise atau outlier buatan pada Dataset Iris. Hal ini dapat mengungkapkan model mana yang lebih tahan terhadap data yang cacat.
2. Eksplorasi Hyperparameter: Melakukan Grid Search atau optimasi hyperparameter (misalnya K dari 1 hingga 25 untuk KNN, dan kombinasi Kernel/C/Gamma untuk SVM) untuk melihat apakah ada konfigurasi yang menghasilkan perbedaan kinerja pada data Iris yang dipecah dalam rasio pembagian data yang berbeda.
3. Perbandingan Komputasi: Mengukur secara eksplisit waktu komputasi pelatihan dan pengujian untuk kedua model pada data yang lebih besar untuk mendapatkan perbandingan efisiensi yang lebih bermakna.

E. DAFTAR PUSTAKA

- Bintoro, P., Ratnasari, W., Wihardjo, E., Putri, I. P., & Asari, A. (2024). Pengantar machine learning. Pustaka Timur.
- Desiani, A., Uteh, C. M., & Rayani, I. (2019). Penerapan metode Support Vector Machine (SVM) dalam klasifikasi bunga iris. *Jurnal Gamma*, 10(1), 50–65.
- Dinata, R. K., & Hasdyna, N. (2020). Machine learning: Panduan memahami data science, supervised learning, unsupervised learning dan reinforcement learning. UNIMAL Press.
- Elisawati, Wijayanti, R. R., & Abdurrasyid. (2023). Machine learning. PT Global Eksekutif Teknologi.
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annual Eugenics*, 7(2), 179–188.
- Géron, A. (2019). Hands-on machine learning with Scikit-Learn, Keras & TensorFlow (2nd ed.). O'Reilly Media.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. The MIT Press.
- Hamonangan, R., Sari, R. K., Anwar, S., & Hartati, T. (2024). Klasifikasi algoritma KNN dalam menentukan penerima bantuan langsung tunai. *Jurnal Ilmiah Sistem Informasi*, 6(1), 198–204.



- Isman, Ahmad, A., & Latief, A. (2021). Perbandingan metode KNN dan LBPH pada klasifikasi daun herbal. *Jurnal RESTI (Rekayasa Sist. dan Teknol. Informasi)*, 5(3), 557–564.
- Rahman, Z., Fatah, Z., & Prasetyo, J. D. (2024). Klasifikasi spesies bunga iris menggunakan algoritma klasifikasi KNN di RapidMiner. *Jurnal Ilmiah Multidisiplin Ilmu*, 1(6), 60–66.
- Ridho, W. A., Hakim, A. R., & Wuryandari, T. (2023). Perbandingan metode K-Nearest Neighbor dan Support Vector Machines pada status penerimaan bantuan dari pemerintah. *Jurnal Gaussian*, 12(3), 372–381.
- Susanto, B., & Wijaya, C. (2023). Implementasi algoritma Naïve Bayes classifier dan K-nearest neighbor untuk klasifikasi penyakit ginjal kronik. *JATISI*, 10(2), 1502–1510.
- Syahputra, A., Riansyah, R., Aptanta, D. A., Farhan, M., & Furqan, M. (2025). Klasifikasi jenis bunga iris menggunakan algoritma klasifikasi tradisional. *Jurnal ilmiah Sistem Informasi dan Ilmu Komputer*, 4(2), 112–120.
- Ujianto, N. T., Gunawan, Fadillah, H., Fanti, A. P., Saputra, A. D., & Ramadhan, I. G. (2025). Penerapan algoritma K-Nearest Neighbors (KNN) untuk klasifikasi citra medis. *IT-Explore: Jurnal Penerapan Teknologi Informasi dan Komunikasi*, 4(1), 33–43.
- Yana, Y. E., & Nafi'iyah, N. (2021). Klasifikasi jenis pisang berdasarkan fitur warna, tekstur, bentuk citra menggunakan SVM dan KNN. *Research Journal of Computer Information System and Technology Management*, 4(1), 28–35. Suparno. 2000. *Langkah-langkah Penulisan Artikel Ilmiah* dalam Saukah, Ali dan Waseso, M.G. 2000. *Menulis Artikel untuk Jurnal Ilmiah*. Malang: UM Press.