

PERBANDINGAN KINERJA ALGORITMA SVM, DECISION TREE, DAN NAIVE BAYES UNTUK KLASIFIKASI DAN PENGELOMPOKAN SPESIES IRIS

Riski Ronaldo¹, Yusuf Kurnia²

^{1,2}Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas
Buddhi Dharma

Email: Ronaldotarigan41@gmail.com

Abstrak

Kemampuan untuk mengelola dan menganalisis data menjadi aspek penting dalam era digital saat ini. Klasifikasi merupakan salah satu metode yang memungkinkan pengelompokan objek berdasarkan pola tertentu. Penelitian ini bertujuan membandingkan kinerja tiga algoritma klasifikasi, yaitu *Support Vector Machine (SVM)*, *Decision Tree*, dan *Naive Bayes*, dalam mengklasifikasikan spesies bunga *Iris*. Dataset *Iris* dipilih karena karakteristiknya yang sering digunakan dalam penelitian klasifikasi, mencakup tiga spesies utama: *Iris setosa*, *Iris versicolor*, dan *Iris virginica*. Penelitian ini menggunakan metodologi berbasis pendekatan terstruktur yang mencakup pemahaman masalah, eksplorasi data, pemrosesan, pemodelan, evaluasi, dan penerapan. Implementasi algoritma dilakukan menggunakan *Python* dengan pustaka *scikit-learn*, sedangkan aplikasi berbasis web dikembangkan menggunakan *framework PHP* dan *Laravel* untuk mendukung pengujian hasil model. Hasil penelitian menunjukkan bahwa *SVM* memiliki kinerja paling unggul dibandingkan dengan *Decision Tree* dan *Naive Bayes*, terutama dalam hal *akurasi*, *presisi*, *recall*, dan *F1-score*. Aplikasi berbasis web yang dibangun mampu memproses gambar bunga yang diunggah untuk menghasilkan prediksi spesies dengan tingkat keakuratan yang tinggi. Temuan ini menunjukkan potensi integrasi algoritma pembelajaran mesin dengan aplikasi berbasis web sebagai solusi praktis dalam pengelompokan dan klasifikasi data secara efisien.

Kata kunci: Klasifikasi, *Machine Learning*, *Spesies Iris*, *Support Vector Machine*.

Pendahuluan

Di era digital dan virtual saat ini, kemampuan untuk mengelola dan menganalisis data menjadi sangat penting. Salah satu metode yang sering digunakan dalam analisis data adalah klasifikasi, yang memungkinkan pengelompokan data berdasarkan pola tertentu (Hanif & Muntiari, 2024). Pengelompokan data telah menjadi teknik yang esensial dalam berbagai bidang, termasuk biologi, medis, dan teknologi (Ramadani et al., 2024). Klasifikasi, sebagai bagian dari *Machine Learning*, bertujuan untuk memprediksi kategori atau label dari data baru berdasarkan model yang telah dilatih menggunakan data

berlabel (Hoffman, n.d.). Salah satu contoh klasik adalah klasifikasi spesies bunga *Iris*, yang dapat diidentifikasi berdasarkan karakteristik fisik seperti panjang dan lebar *sepal* serta panjang dan lebar *petal*. Dataset *Iris* sering digunakan sebagai tolok ukur dalam penelitian klasifikasi karena kesederhanaannya namun tetap menawarkan tantangan analisis yang menarik (Yehya Hussien, 2022). Penelitian ini berfokus pada perbandingan tiga algoritma klasifikasi utama: *Support Vector Machine* (SVM), *Decision Tree* (DT), dan *Naive Bayes* (NB). SVM bekerja dengan mencari *hyperplane* terbaik untuk memisahkan data dengan margin yang maksimal (Start-up, 2021). DT membangun model prediksi dalam bentuk struktur pohon keputusan berdasarkan atribut data, sedangkan NB merupakan algoritma probabilistik berbasis teori Bayes yang mengasumsikan independensi antar fitur. Penelitian ini bertujuan untuk mengevaluasi kinerja ketiga algoritma tersebut dalam klasifikasi spesies *Iris*. Evaluasi dilakukan menggunakan metrik seperti akurasi, presisi, *recall*, dan *F1-score*. Selain itu, penelitian ini juga mendokumentasikan proses pengolahan data menggunakan metodologi *CRISP-DM* (Cross Industry Standard Process for Data Mining) (Schröer et al., 2021), aplikasi berbasis web yang dikembangkan menggunakan bahasa pemrograman PHP dan framework Laravel bertujuan untuk mengintegrasikan hasil analisis ke dalam sistem yang lebih praktis. (Guide et al., 2022) Manfaat dari penelitian ini adalah memberikan wawasan baru mengenai kinerja algoritma SVM, DT, dan NB dalam klasifikasi, serta menjadi referensi praktis untuk tugas serupa. Selain itu, penelitian ini dapat memberikan panduan implementasi algoritma *Machine Learning* menggunakan Python, sekaligus mendukung pengembangan aplikasi berbasis web untuk klasifikasi. Ruang lingkup penelitian meliputi pemilihan dataset *Iris* (Salsabila et al., 2021), implementasi algoritma SVM, DT, dan NB, evaluasi dan analisis hasil, serta pengembangan aplikasi berbasis web sebagai solusi praktis dalam klasifikasi spesies *Iris*. Melalui penelitian ini, diharapkan ditemukan algoritma dengan kinerja terbaik dalam klasifikasi spesies bunga *Iris*, serta potensi penerapan hasil penelitian ini dalam pengembangan teknologi dan aplikasi berbasis data di berbagai bidang industri dan ilmu pengetahuan (Tuntun et al., 2022).

Metodologi

Penelitian ini menggunakan metodologi *CRISP-DM* (*Cross-Industry Standard Process for Data Mining*). Tujuan utama penelitian ini adalah membandingkan kinerja algoritma *Support Vector Machine* (SVM), *Decision Tree* (DT), dan *Naive Bayes* (NB) dalam klasifikasi spesies *Iris* (Mithy et al., 2022). Algoritma dengan performa terbaik selanjutnya diimplementasikan dalam aplikasi berbasis web untuk mendukung kebutuhan sektor botani, agribisnis, dan pendidikan. Data utama yang digunakan berasal dari dataset *Iris* yang tersedia di *UCI Machine Learning Repository* (Farhan & Setiaji, 2023). Dataset ini mencakup fitur panjang dan lebar *sepal* serta *petal* untuk tiga spesies *Iris* (*Setosa*, *Versicolor*, dan *Virginica*). Proses penelitian melibatkan beberapa tahap, yaitu pembersihan, transformasi, dan normalisasi dataset. Data tersebut kemudian dibagi menjadi dua bagian: data latih (70%) dan data uji (30%), untuk memastikan kualitas model yang dihasilkan. Model klasifikasi dibangun dengan menggunakan algoritma SVM, DT, dan NB yang diimplementasikan menggunakan pustaka *Scikit-learn* (Nalatissifa et al., 2021). Setiap algoritma dilatih menggunakan data latih dan diuji pada data uji untuk menghasilkan metrik kinerja seperti akurasi, presisi, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa algoritma SVM dan DT memberikan performa terbaik dalam klasifikasi spesies *Iris* (Nurajijah & Riana, 2019), sementara NB tetap menunjukkan kinerja yang cukup kompetitif. Visualisasi hasil evaluasi juga dilakukan untuk memperkuat analisis yang dilakukan. Implementasi algoritma terbaik ini diharapkan dapat memberikan kontribusi nyata dalam pengembangan teknologi berbasis data pada berbagai sektor terkait.

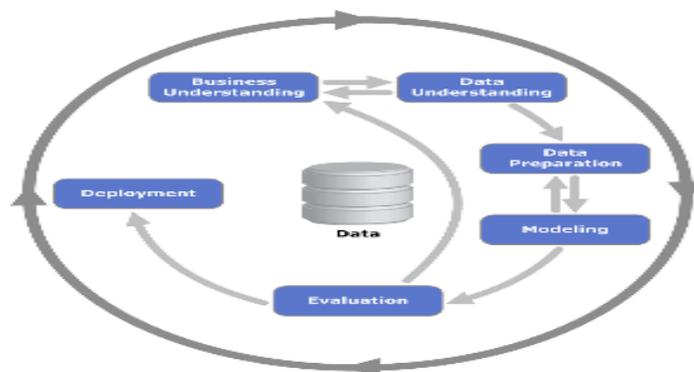


Diagram Metodologi Penelitian

1. Business Understanding:
 - Membandingkan performa *SVM*, *DT*, & *NB* pengelompokan spesies Iris.
 - Menentukan algoritma mana yang memiliki kinerja terbaik berdasarkan metrik akurasi, precision, recall, dan F1-score.
 - Mengimplementasikan metodologi CRISP-DM untuk pengolahan data menggunakan Python dan scikit-learn.
2. Data Understanding
 - Mengumpulkan data.
 - Eksplorasi data awal.
3. Data Preparation:
 - Pembersihan data.
 - Transformasi data.
 - Pembagian data menjadi data latih dan data uji.
4. Modeling:
 - Implementasi algoritma SVM, Decision Tree, dan Naive Bayes.
 - Tuning hyperparameter.
5. Evaluation:
 - Evaluasi kinerja model menggunakan data uji.
 - Analisis hasil evaluasi.
6. Deployment:
 - Aplikasi Analisis Spesies Iris Berbasis Web.

Hasil dan Pembahasan

Penelitian ini membahas perbandingan kinerja tiga algoritma klasifikasi yang banyak digunakan, yaitu *Support Vector Machine* (SVM), *Decision Tree* (DT), dan *Naive Bayes* (NB), dengan menggunakan dataset *Iris* dari *UCI Machine Learning Repository*. Dataset tersebut terdiri dari 150 sampel dengan empat fitur utama, yaitu panjang dan lebar *sepal* serta panjang dan lebar *petal*, yang diklasifikasikan ke dalam tiga kelas: *Setosa*, *Versicolor*, dan *Virginica*. Data dibagi dengan rasio 70:30 untuk pelatihan dan pengujian menggunakan fungsi

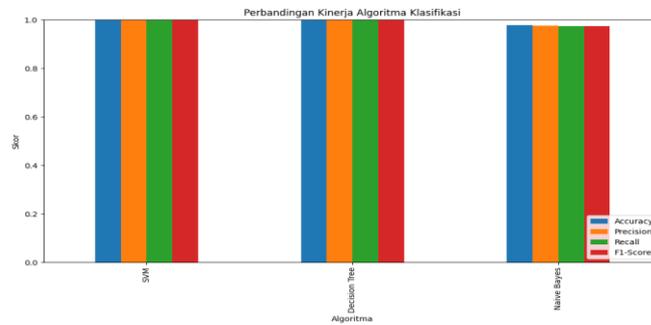
train_test_split dari pustaka *Scikit-learn* (Dani & Ginting, 2024). Evaluasi kinerja algoritma dilakukan menggunakan metrik akurasi, presisi, *recall*, dan *F1-score*. Berdasarkan hasil evaluasi, algoritma *Support Vector Machine* (SVM) dan *Decision Tree* (DT) berhasil mencapai performa sempurna dengan skor 1.000 pada seluruh metrik yang digunakan. Sebaliknya, algoritma *Naive Bayes* (NB) mencatatkan skor akurasi sebesar 0.978. Meskipun sedikit lebih rendah dibandingkan SVM dan DT, algoritma NB tetap menunjukkan kinerja yang sangat baik dan layak diaplikasikan dalam tugas klasifikasi yang serupa (Santoso et al., 2023). Temuan ini menunjukkan bahwa masing-masing algoritma memiliki potensi unggul dalam menyelesaikan permasalahan klasifikasi sesuai dengan karakteristik data yang digunakan (Asniar et al., 2022).

Tabel 1 Hasil Perbandingan

Algoritma	Akurasi	Presisi	Recall	F1-Score
SVM	1.000	1.000	1.000	1.000
Decision Tree	1.000	1.000	1.000	1.000
Naive Bayes	0,978	0,976	0,974	0,974

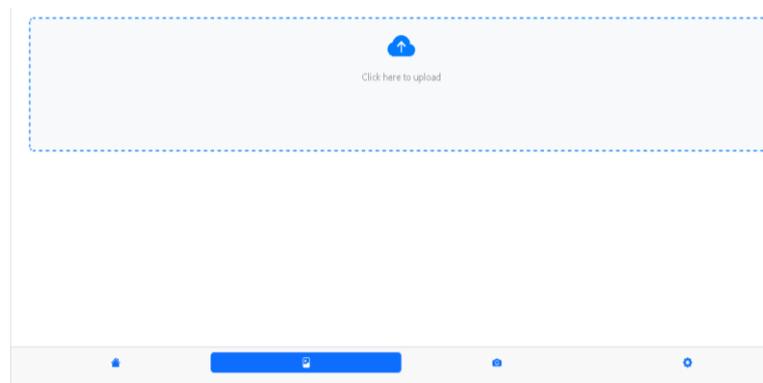
Support Vector Machine (SVM): SVM menunjukkan kinerja terbaik dengan akurasi sempurna. Algoritma ini sangat efektif untuk dataset yang memiliki jumlah fitur terbatas. Namun, SVM memiliki kelemahan dalam hal waktu komputasi yang cenderung lebih lama, terutama ketika diterapkan pada dataset yang besar. *Decision Tree*: Algoritma *Decision Tree* memberikan performa yang setara dengan SVM. Salah satu keunggulannya adalah kemudahan interpretasi melalui visualisasi struktur pohon keputusan. Namun, algoritma ini memiliki kekurangan berupa kerentanan terhadap *overfitting* saat digunakan pada dataset dengan tingkat kompleksitas yang tinggi. *Naive Bayes*: *Naive Bayes* menunjukkan kinerja yang baik dengan keunggulan efisiensi komputasi. Akan tetapi, asumsi dasar algoritma ini mengenai independensi antar fitur menjadi kelemahan

signifikan ketika diterapkan pada dataset yang memiliki tingkat korelasi fitur yang tinggi.

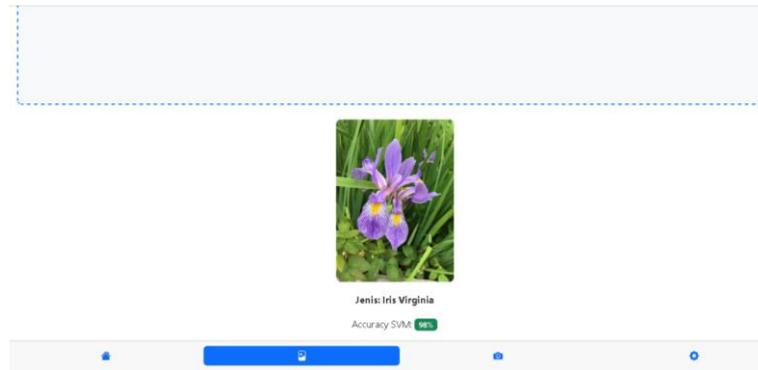


Gambar 1 1 Hasil Evaluasi

Visualisasi yang disajikan memperlihatkan perbandingan kinerja algoritma berdasarkan metrik akurasi, presisi, *recall*, dan *F1-score*, yang memberikan gambaran jelas mengenai performa masing-masing algoritma dalam tugas klasifikasi. Aplikasi berbasis web yang dikembangkan menggunakan *framework Laravel* pada sisi *backend* serta *JavaScript* dan *CSS* untuk *frontend*. Aplikasi ini dirancang agar pengguna dapat mengunggah gambar bunga Iris atau menggunakan kamera perangkat untuk melakukan prediksi spesies.



Aplikasi berhasil memprediksi spesies bunga Iris berdasarkan gambar yang diunggah dengan tingkat akurasi yang ditampilkan secara *real-time*. Hasil pengujian menunjukkan bahwa aplikasi ini berpotensi menjadi alat bantu interaktif yang efektif untuk mendukung proses klasifikasi spesies bunga secara akurat.



Gambar 1 2 Hasil Pengujian Aplikasi

Simpulan

Telah ditetapkan dengan memberikan evaluasi komprehensif terhadap algoritma klasifikasi dan penerapannya pada dataset spesies Iris. Hasil penelitian menunjukkan bahwa algoritma *Support Vector Machine (SVM)* dan Decision Tree memberikan kinerja sempurna, sementara Naive Bayes meskipun sedikit lebih rendah, tetap menunjukkan kinerja yang sangat baik. Penelitian ini juga berhasil mengimplementasikan metodologi *CRISP-DM* untuk pengolahan data, serta mengembangkan aplikasi berbasis web yang dapat memprediksi spesies bunga *Iris* berdasarkan gambar. Aplikasi ini menunjukkan penerapan praktis dari model machine learning dalam platform yang ramah pengguna. Secara keseluruhan, penelitian ini memberikan kontribusi dalam pemahaman perbandingan algoritma *klasifikasi* dan pengembangan aplikasi berbasis web.

Daftar Pustaka

- Asniar, Maulidevi, N. U. & Surendro, K. (2022). SMOTE-LOF for noise identification in imbalanced data classification. *Journal of King Saud University - Computer and Information Sciences*, 34(6), 3413–3423. <https://doi.org/10.1016/j.jksuci.2021.01.014>
- Dani, Y. & Ginting, M. A. (2024). Comparison of Iris dataset classification with Gaussian naïve Bayes and decision tree algorithms. *International Journal of Electrical and Computer Engineering*, 14(2), 1959–1968. <https://doi.org/10.11591/ijece.v14i2.pp1959-1968>
- Farhan, N. M. & Setiaji, B. (2023). Indonesian Journal of Computer Science. *Indonesian Journal of Computer Science*, 12(2), 284–301. <http://ijcs.stmikindonesia.ac.id/ijcs/index.php/ijcs/article/view/3135>
- Guide, P., Sciences, T. & Nadu, T. (2022). *Decision Tree Over Support Vector Machine for Better Accuracy in Identifying the Problem Based on the Iris Flower*. 0. <https://doi.org/10.3233/APC220028>

- Hanif, H. & Muntiar, N. R. (2024). *PENERAPAN ALGORITMA DECISION TREE, SVM, NAÏVE BAYES DALAM DETEKSI STUNTING PADA BALITA*. 8(1), 105–109.
- Hoffman, D. W. (n.d.). *No 主観的健康感を中心とした在宅高齢者における健康関連標に関する共分散構造分析* Titleasdas.
- Mithy, S. A., Hossain, S., Akter, S., Honey, U. & Sogir, S. B. (2022). Classification of Iris Flower Dataset using Different Algorithms. *International Journal of Scientific Research in Research Paper: Mathematical and Statistical Sciences*, 9(6), 1–10. <https://doi.org/10.26438/ijrmss/v9i6.110>
- Nalattissifa, H., Gata, W., Diantika, S. & Nisa, K. (2021). Perbandingan Kinerja Algoritma Klasifikasi Naive Bayes, Support Vector Machine (SVM), dan Random Forest untuk Prediksi Ketidakhadiran di Tempat Kerja. *Jurnal Informatika Universitas Pamulang*, 5(4), 578. <https://doi.org/10.32493/informatika.v5i4.7575>
- Nurajijah, N. & Riana, D. (2019). Algoritma Naïve Bayes, Decision Tree, dan SVM untuk Klasifikasi Persetujuan Pembiayaan Nasabah Koperasi Syariah. *Jurnal Teknologi Dan Sistem Komputer*, 7(2), 77–82. <https://doi.org/10.14710/jtsiskom.7.2.2019.77-82>
- Ramadani, N. C., Tahyudin, I. & Shouni Barkah, A. (2024). Perbandingan Algoritma Support Vector Machine, Decision Tree, dan Logistic Regression Pada Analisis Sentimen Ulasan Aplikasi Netflix. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 10(2), 110–117. <https://teknosi.fti.unand.ac.id/index.php/teknosi/article/view/2746>
- Salsabila, A., Yunita, R. & Rozikin, C. (2021). Identifikasi Citra Jenis Bunga menggunakan Algoritma KNN dengan Ekstraksi Warna HSV dan Tekstur GLCM. *Technomedia Journal*, 6(1), 124–137. <https://doi.org/10.33050/tmj.v6i1.1667>
- Santoso, M. S., Suryadi, J. J., Marchellino, K., Nabiilah, G. Z. & Rojali. (2023). A Comparative Analysis of Decision Tree and Support Vector Machine on Suicide Ideation Detection. *Procedia Computer Science*, 227, 518–523. <https://doi.org/10.1016/j.procs.2023.10.553>
- Schröer, C., Kruse, F. & Gómez, J. M. (2021). A systematic literature review on applying CRISP-DM process model. *Procedia Computer Science*, 181(2019), 526–534. <https://doi.org/10.1016/j.procs.2021.01.199>
- Start-up, P. K. (2021). *Analisis Perbandingan Algoritma Decision Tree, kNN, dan Naive Bayes untuk Analisis Perbandingan Algoritma Decision Tree, kNN, dan Naive Bayes untuk Prediksi Kesuksesan Start-up*. September. <https://doi.org/10.14421/jiska.2021.6.3.178-188>
- Tuntun, R., Kusriani, K. & Kusnawi, K. (2022). Analisis Perbandingan Kinerja Algoritma Klasifikasi dengan Menggunakan Metode K-Fold Cross Validation. *Jurnal Media Informatika Budidarma*, 6(4), 2111. <https://doi.org/10.30865/mib.v6i4.4681>
- Yehya Hussien, A. (2022). Flower Species Recognition Using Machine Learning Classifiers. *Academic Journal of Nawroz University*, 11(4), 469–475. <https://doi.org/10.25007/ajnu.v11n4a1636>