

Analisis Distribusi Kinerja SVM dan KNN Berdasarkan Rata Rata Simpangan Baku dan Stabilitas

Fransciko Pritama*¹, Ekat Rueh Daya Leluni², Yovita³, Jadiaman Parhusip⁴

Universitas Palangkaraya, Jl. Yos Sudarso Palangka Kec. Jekan Raya Kota Palangka Raya 74874

*¹email: fransciko.pritama02@gmail.com

²email: ekatruehdayaleluni22@gmail.com

³email: yopitaipit1301@gmail.com

⁴email: parhusip.jadiaman@it.upr.ac.id

(Naskah diterima: 18 Nopember 2024; Naskah direvisi: 7 Desember 2024; Naskah diterbitkan: 10 Desember 2024)

ABSTRAK – Di bidang pembelajaran mesin, dua algoritma yang umum digunakan untuk tugas klasifikasi adalah Support Vector Machine (SVM) dan K-Nearest Neighbor (KNN). Kedua algoritma tersebut dapat digunakan dalam berbagai aplikasi, namun memilih algoritma terbaik memerlukan analisis rinci mengenai kinerjanya dalam skenario yang berbeda. Tujuan artikel ini adalah membandingkan performa kedua algoritma menggunakan metrik statistik utama yaitu, Akurasi Rata-Rata, Deviasi Standar, dan Stabilitas Performa. Untuk mengevaluasi algoritma, teknik k-fold cross-validation digunakan pada kumpulan data klasifikasi tertentu. Hal ini memungkinkan Anda mengukur keakuratan model pada rentang data yang berbeda. Berdasarkan hasil penelitian, rata-rata akurasi SVM lebih tinggi dibandingkan dengan ANN, dengan nilai SVM sebesar 0,980 dan nilai ANN sebesar 0,953. Selain itu, deviasi standar kinerja SVM yang lebih rendah (0,043 dibandingkan dengan 0,052 untuk KNN) menunjukkan bahwa kinerja SVM lebih stabil dan konsisten di seluruh konvolusi. Hal ini membuat SVM lebih andal ketika menangani data baru, terutama dataset dengan pola yang kompleks dan sulit dipisahkan. Temuan ini memberikan panduan penting dalam memilih algoritma klasifikasi berdasarkan kebutuhan spesifik aplikasi. SVM lebih disukai ketika stabilitas dan keandalan hasil diperlukan, sedangkan ANN mungkin lebih cocok untuk masalah sederhana dan kumpulan data kecil. Memahami distribusi performa algoritme ini dapat membantu pengguna membuat keputusan yang lebih baik saat menerapkan pembelajaran mesin.

Kata Kunci – Distribusi Kinerja, K-Nearest Neighbors, Simpangan Baku, Support Vector Machine

Analysis of SVM and KNN Performance Distribution Based on Average Standard Deviation and Stability

ABSTRACT – In the field of machine learning, two algorithms commonly used for classification tasks are Support Vector Machine (SVM) and K-Nearest Neighbor (KNN). Both algorithms can be used in various applications, but choosing the best algorithm requires a detailed analysis of its performance in different scenarios. The aim of this article is to compare the performance of both algorithms using key statistical metrics namely, Average Accuracy, Standard Deviation, and Performance Stability. To evaluate the algorithm, the k-fold cross-validation technique was used on a specific classification data set. This allows you to measure model accuracy over different data ranges. Based on the research results, the average accuracy of SVM is higher than that of ANN, with an SVM value of 0.980 and an ANN value of 0.953. Additionally, the lower standard deviation of SVM performance (0.043 compared to 0.052 for KNN) indicates that SVM performance is more stable and consistent across convolutions. This makes SVM more reliable when handling new data, especially datasets with complex patterns that are difficult to separate. These findings provide important guidance in selecting classification algorithms based on application-specific needs. SVM is preferred when stability and reliability of results are required, whereas ANN may be better suited for simple problems and small data sets. Understanding the performance distribution of these algorithms can help users make better decisions when applying machine learning.

Keywords - Support Vector Machine, K-Nearest Neighbors, performance distribution, average, standard deviation

1. PENDAHULUAN

Di era informasi yang terus berkembang, memilih algoritma pembelajaran mesin (*machine learning*) yang tepat merupakan elemen penting dalam pemrosesan data. *Support Vector Machine* (SVM) dan *K-Nearest Neighbor* (KNN) adalah dua algoritma yang banyak digunakan untuk tugas klasifikasi dan regresi karena efektivitas dan fleksibilitasnya. Namun distribusi performa kedua algoritma ini sangat dipengaruhi oleh beberapa karakteristik data, parameter model yang diterapkan, dan teknik pra-pemrosesan yang digunakan [1]. Hal ini menyoroti pentingnya memiliki pemahaman mendalam tentang kinerja algoritma dalam berbagai skenario.

Pendekatan yang umum digunakan atau diterapkan untuk mengevaluasi algoritma mencakup analisis distribusi kinerja menggunakan metrik statistik seperti akurasi rata-rata, deviasi standar, dan stabilitas model. Akurasi rata-rata mencerminkan kemampuan algoritme untuk mengklasifikasikan seluruh data, sedangkan deviasi standar mengukur konsistensi kinerja di berbagai subkumpulan data. Stabilitas model mewakili sensitivitas algoritma terhadap perubahan kecil dalam kumpulan data. Ini merupakan aspek penting dalam aplikasi dunia nyata.

Penelitian sebelumnya oleh Saputra dan Putri (2023) [2] menunjukkan bahwa parameter seperti kernel pada SVM dan nilai K pada KNN mempunyai pengaruh yang signifikan terhadap hasil klasifikasi, terutama untuk dataset dengan pola yang kompleks. Selain itu, Rahman (2022) [3], dalam studinya tentang kinerja algoritma pembelajaran mesin, menyoroti peran teknik pra-pemrosesan seperti regularisasi dan pemilihan fitur dalam meningkatkan akurasi dan stabilitas model. Temuan tersebut memberikan dasar penting untuk mengkaji kinerja kedua algoritma dalam penelitian ini.

Penelitian ini bertujuan untuk menganalisis distribusi kinerja algoritma SVM dan ANN dengan mengamati tiga metrik utama: akurasi rata-rata, standar deviasi, dan stabilitas model. Memahami pola kinerja kedua algoritma ini dapat membantu pengguna membuat keputusan yang lebih tepat ketika memilih algoritma yang tepat untuk kebutuhan spesifik mereka.

2. TINJAUAN PUSTAKA

2.1. *Support Vector Machine* (SVM)

Support Vector Machine (SVM) adalah jenis algoritma klasifikasi yang populer digunakan di banyak aplikasi pembelajaran mesin. SVM bekerja untuk menemukan hyperplane yang memisahkan data dengan jarak maksimum, sehingga memperluas jarak antara berbagai kelas [4]. SVM mampu

menangani data dengan banyak fitur dan sering diandalkan untuk tugas pengenalan pola serta analisis data teks [5]. Salah satu keunggulan utama SVM adalah efisiensinya dalam mengelola data non-linear dengan menggunakan model kernel, yang memungkinkan pemisahan data dalam sebuah ruang berdimensi lebih tinggi [6].

2.2. *K-Nearest Neighbors* (KNN)

K-Nearest Neighbors (KNN) merupakan algoritma pembelajaran yang dianggap berbasis instance yang mengklasifikasikan sebuah data baru berdasarkan ketersambungan dengan data yang sudah ada [7]. KNN banyak digunakan dalam tugas klasifikasi dan regresi yang sederhana, tetapi kinerjanya sangat dipengaruhi oleh aktifitas pemilihan jumlah tetangga terdekat (K) dan metode pengukuran jarak [8]. Meskipun tergolong pada tingkat sederhana, KNN bisa menghasilkan performa baik pada dataset dengan distribusi yang jelas dan menyeluruh. Salah satu kelemahan KNN adalah waktu komputasi yang tinggi pada data besar, karena penghitungan jarak antara titik data dilakukan secara langsung untuk setiap prediksi [9].

2.3. Evaluasi Kinerja Model dalam Pembelajaran Mesin

Evaluasi kinerja model dalam pembelajaran mesin merupakan hal penting untuk menilai apakah model yang diterapkan dapat memberikan hasil yang terbaik dalam berbagai situasi. Salah satu metode evaluasi yang banyak digunakan adalah *k-fold cross-validation*, yang membagi dataset menjadi *k* subset dan melatih model pada *k-1* subset untuk diuji pada subset lainnya. Metode ini bisa memberikan gambaran lebih tepat mengenai kemampuan model dalam menggeneralisasi data baru [10]. Proses evaluasi ini menghasilkan metrik seperti akurasi, rata-rata, simpangan baku, dan stabilitas, yang semuanya dapat digunakan untuk menilai kualitas keseluruhan modelnya. Rata-rata menunjukkan gambaran umum kinerja model, sedangkan simpangan baku memberikan informasi tentang variasi hasil di setiap iterasi eksperimen [11].

2.4. Perbandingan SVM dan KNN dalam Tugas Klasifikasi

Sejumlah kajian telah mengevaluasi performa SVM dan KNN dalam tugas klasifikasi dengan berbagai tipe dataset. SVM sering menunjukkan hasil yang lebih baik dalam hal akurasi dan stabilitas, terutama pada dataset yang memiliki banyak fitur dan bersifat non-linear [12]. Di sisi lain, KNN lebih mudah dan umumnya lebih cepat dalam mengolah data yang lebih kecil dan lebih seragam. Perbandingan ini mengindikasikan bahwa pemilihan

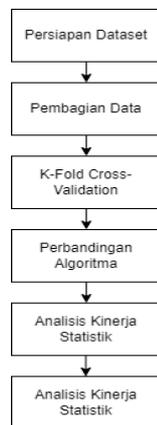
algoritma sangat tergantung pada karakteristik dataset dan tujuan analisis [13].

2.5. Stabilitas Kinerja Model Pembelajaran Mesin

Stabilitas performa model sangat penting untuk menilai seberapa handal model dapat memberikan hasil yang konsisten dalam berbagai kondisi. Model dengan stabilitas tinggi menghasilkan keluaran yang konsisten meskipun diterapkan pada berbagai subset dari data. Sebaliknya, model dengan fluktuasi hasil yang tinggi menunjukkan ketergantungan besar pada data pelatihan tertentu, yang dapat menjadi masalah saat menguji model dalam situasi nyata [14]. Stabilitas menjadi salah satu faktor penting yang perlu diperhatikan saat memilih model klasifikasi [15].

3. METODE PENELITIAN

Penelitian ini bertujuan untuk menganalisis distribusi kinerja dua algoritma pembelajaran mesin, *Support Vector Machine* (SVM) dan *K-Nearest Neighbors* (KNN), menggunakan metrik statistik penting seperti akurasi rata-rata, standar deviasi, dan stabilitas.



Gambar 1. Metode Penelitian

Metode penelitian ini dibagi menjadi beberapa langkah seperti pada gambar 1.

3.1. Persiapan Dataset

Dataset Iris dipilih karena berisi fitur-fitur yang relevan dengan tugas klasifikasi dengan dimensi fitur berbeda. Dataset dibagi menjadi dua bagian utama, training set digunakan untuk melatih model dan set pengujian digunakan untuk mengukur performa model pada data yang belum pernah dilihat sebelumnya.

3.2. Pembagian Data

Data dibagi menjadi dua set utama antara lain set pelatihan dan set pengujian. Pemisahan data dilakukan dengan cara teknik split pelatihan dan

pengujian dengan rasio set pelatihan sebesar 70% dan pengujian sebesar 30%. Teknik ini digunakan untuk memastikan bahwa model yang Anda bangun dapat dievaluasi pada data yang belum pernah dilihat sebelumnya.

3.3. K-fold Cross-Validation

Untuk mengevaluasi model yang dibuat secara lebih obyektif, penelitian ini menggunakan metode k-fold cross validation. Validasi silang dilakukan dengan membagi kumpulan data menjadi k subset, dan model dilatih pada subset k-1 dan diuji pada subset yang tersisa. Proses ini diulangi sebanyak k kali, sehingga setiap subset mempunyai kesempatan untuk digunakan sebagai data uji. Nilai k yang digunakan dalam penelitian ini adalah 10 yang dikenal dengan istilah 10-fold cross validation. Evaluasi dilakukan dengan mengukur keakuratan setiap lipatan.

3.4. Perbandingan Algoritma

Dua algoritma digunakan dalam penelitian ini yaitu SVM dan KNN. *Support Vector Machine* (SVM) menggunakan kernel linear untuk klasifikasi dan parameter default lainnya. *K-Nearest Neighbor* (KNN) dimana parameter K dipilih melalui eksperimen untuk menemukan sebuah nilai K terbaik yang dapat menghasilkan akurasi tertinggi. Model SVM dan KNN ini dilatih menggunakan training set pada setiap iterasi cross-validation dan dievaluasi dengan mengukur akurasi pada test set.

3.5. Analisis Kinerja Statistik

Analisis Kinerja Statistik Setelah validasi silang selesai, hasil akurasi untuk setiap lipatan digunakan untuk analisis distribusi kinerja. Statistik yang dihitung adalah rata-rata untuk Mengukur performa model secara keseluruhan. Kemudian teknik deviasi standar untuk Mengevaluasi variasi kinerja model. Dan stabilitas dihitung dari konsistensi hasil kinerja di kali lipat.

3.6. Perbandingan Kinerja Distribusi

Setelah validasi silang selesai, hasil setiap lipatan kedua algoritma digabungkan untuk analisis distribusi. Rata-rata dan deviasi standar digunakan untuk mengevaluasi kinerja setiap algoritma. Perbandingan kemudian dibuat antara SVM dan KNN dengan memeriksa seberapa baik dan seberapa konsisten kinerja kedua model pada subkumpulan data yang berbeda.

4. HASIL DAN PEMBAHASAN

4.1. Hasil Cross-Validation

Tabel 1 menyajikan akurasi per fold untuk kedua algoritma. Algoritma yang dimaksud adalah SVM dan KNN.

Tabel 1. Cross-Validation

Fold	Akurasi SVM	Akurasi KNN
Fold 1	1.00	1.00
Fold 2	1.00	1.00
Fold 3	1.00	1.00
Fold 4	1.00	0.93
Fold 5	0.93	0.87
Fold 6	0.87	0.87
Fold 7	1.00	1.00
Fold 8	1.00	1.00
Fold 9	1.00	0.93
Fold 10	1.00	0.93

Pada Tabel 1 diperoleh hasil cross-validation untuk algoritma SVM dan KNN. Hasil tersebut menunjukkan nilai yang bervariasi pada setiap fold. Cross-validation membagi data menjadi 10 bagian (fold), di mana pada setiap fold secara bergantian digunakan sebagai data uji, sementara sisa data lainnya digunakan untuk melatih model. Pada setiap iterasi, model akan dihitung akurasinya berdasarkan kemampuannya mengklasifikasikan data uji.

4.2. Rata-Rata Akurasi

Rata-rata akurasi dihitung untuk mengevaluasi kinerja keseluruhan model.

Rumus rata-rata:

$$\mu = \frac{\sum xi}{n}$$

Rata-Rata Akurasi SVM :

$$\mu_{SVM} = \frac{9.80}{10} = 0.980$$

Rata-Rata Akurasi KNN :

$$\mu_{KNN} = \frac{9.53}{10} = 0.953$$

Rata-rata Akurasi : SVM lebih unggul (0.980) dibandingkan KNN (0.953).

4.3. Simpangan Baku

Simpangan baku dihitung untuk mengukur konsistensi kinerja model di setiap fold.

Rumus Simpangan Baku :

$$\sigma = \sqrt{\frac{\sum (xi - \mu)^2}{n}}$$

Perhitungan untuk SVM :

- Selisih dari rata-rata :
[0.02,0.02,0.02,0.02,-0.05,-0.11,0.02,0.02,0.02,0.02]
- Kuadrat hasil selisih :
[0.0004,0.0004,0.0004,0.0004,0.0025,0.0121,0.0004,0.0004,0.0004,0.0004]

- Hitung rata-rata kuadrat selisih (varians) :
$$variance_{SVM} = \frac{0.0004 \cdot 8 + 0.0025 + 0.0121}{10} = 0.00185$$
- Akar Varians
$$\sigma_{SVM} = \sqrt{0.00185} = 0.043$$

Perhitungan Untuk KNN

- Selisih dari rata-rata :
[0.047,0.047,0.047,-0.023,-0.083,-0.083,0.047,0.047,-0.023,-0.023]
- Kuadrat hasil selisih :
[0.0022,0.0022,0.0022,0.0005,0.0069,0.0069,0.0022,0.0022,0.0005,0.0005]
- Hitung rata-rata kuadrat selisih (varians) :
$$variance_{KNN} = \frac{0.0022 \cdot 8 + 0.0005 + 0.0069 \cdot 2}{10} = 0.0027$$
- Akar Varians
$$\sigma_{KNN} = \sqrt{0.0027} = 0.052$$

4.4. Stabilitas Kinerja

SVM menunjukkan simpangan baku lebih kecil (0.043) dibandingkan KNN (0.052). Ini menandakan bahwa hasil SVM lebih konsisten antar fold.

5. SIMPULAN

Studi ini menunjukkan bahwa *Support Vector Machine* (SVM) memiliki rata-rata akurasi yang lebih tinggi dibanding *K-Nearest Neighbors* (KNN) dalam tugas klasifikasi. SVM mendapatkan nilai rata-rata akurasi 0.980, sedangkan KNN mencapai 0.953, yang menunjukkan SVM lebih baik dalam memisahkan kelas di dataset yang dianalisis. Ini berarti bahwa SVM lebih efisien dalam menangani pola-pola rumit dalam analisis data. Selain akurasi, stabilitas kinerja algoritma juga penting. SVM memiliki simpangan baku 0.043, yang lebih kecil dibanding KNN dengan simpangan baku 0.052. Angka ini menunjukkan bahwa kinerja SVM lebih konsisten di antara fold selama proses k-fold cross-validation, sehingga lebih bisa diandalkan untuk diterapkan pada data baru. Dalam hal distribusi kinerja, SVM menunjukkan hasil yang lebih konsisten dibanding KNN. Kinerja yang stabil ini membuat SVM lebih tepat digunakan dalam aplikasi klasifikasi yang memerlukan hasil yang dapat diprediksi dengan akurat. Di sisi lain, KNN lebih mudah terpengaruh oleh pemilihan parameter (K) dan variasi dalam data, yang dapat mengakibatkan perubahan dalam hasil klasifikasi. Secara keseluruhan, analisis distribusi kinerja yang meliputi rata-rata, simpangan baku, dan stabilitas kinerja menekankan keunggulan SVM dibanding KNN pada dataset yang dianalisis. Dengan demikian, pemilihan algoritma pada klasifikasi perlu

mempertimbangkan kebutuhan spesifik aplikasi, terutama terkait akurasi dan stabilitas hasil.

DAFTAR PUSTAKA

- [1] Sisfokom. (2022). Analisis Kinerja Algoritma Klasifikasi dengan Teknik Preprocessing pada Dataset Bahasa Indonesia. *Jurnal Sistem Informasi dan Komputer*, 11(1), 25-31.
- [2] Saputra, R., & Putri, A. (2023). Pengaruh Parameter dan Teknik Praproses pada Algoritma Klasifikasi. *Jurnal Informatika dan Sains Data*, 11(2), 45-53.
- [3] Rahman, A. (2022). Analisis Performansi Algoritma Pembelajaran Mesin Berdasarkan Variasi Parameter. *Jurnal Teknologi Informasi dan Komputer*, 9(3), 78-89.
- [4] Putri, R. (2020). Klasifikasi data menggunakan algoritma SVM dengan kernel linear untuk pengenalan pola. *Jurnal Teknik Informatika*, 9(2), 85-97.
- [5] Ningsih, R., & Putra, Y. (2021). Support vector machine untuk klasifikasi teks: Studi kasus analisis sentimen. *Jurnal Komputasi dan Informatika*, 11(4), 221-234.
- [6] Hidayat, M., & Supriyanto, S. (2022). Analisis penggunaan kernel pada SVM untuk klasifikasi data besar. *Jurnal Teknologi dan Sistem Informasi*, 15(2), 76-89.
- [7] Wahyu, A. (2019). Implementasi algoritma KNN untuk klasifikasi data sensor. *Jurnal Teknologi dan Rekayasa*, 12(1), 31-44. *The Oxford Dictionary of Computing*, 5th ed. Oxford: Oxford University Press, 2003.
- [8] Nurrahmawati, A., & Sulistiyo, D. (2020). Pengaruh pemilihan nilai K pada algoritma K-Nearest Neighbors untuk prediksi harga rumah. *Jurnal Ilmu Data dan Teknologi*, 6(3), 179-191.
- [9] Pradipta, A. (2021). Kelemahan algoritma KNN dalam pengolahan data besar. *Jurnal Data Mining dan Analitik*, 13(1), 54-67.
- [10] Setiawan, D., & Pratama, S. (2020). Penerapan k-fold cross-validation untuk evaluasi model pembelajaran mesin. *Jurnal Sistem Informasi*, 7(1), 42-53.
- [11] Gunawan, E., & Setiawan, A. (2021). Evaluasi model pembelajaran mesin menggunakan k-fold cross-validation dan perhitungan statistik. *Jurnal Ilmu Komputer*, 8(1), 45-58.
- [12] Sari, D., & Agustina, I. (2019). Perbandingan kinerja algoritma SVM dan KNN dalam klasifikasi citra digital. *Jurnal Komputasi dan Teknologi*, 5(2), 89-101.
- [13] Pratiwi, L., & Nugroho, P. (2018). Perbandingan Kinerja Algoritma K-Nearest Neighbors dan Support Vector Machine dalam Klasifikasi Data. *Jurnal Teknologi Informasi*, 12(2), 105-114.
- [14] Handayani, D., & Pramudito, D. (2020). Stabilitas model dalam klasifikasi: Pengaruh simpangan baku terhadap kinerja model. *Jurnal Pembelajaran Mesin Indonesia*, 4(3), 102-112.
- [15] Hafidhoh, N., Santosa, P. I., & Sari, R. P. (2024). Predictive Maintenance System menggunakan Algoritma Pembelajaran Mesin. *Jurnal Masyarakat Informatika*, 15(1), 59-61.