

ANALISIS PERBANDINGAN AKURASI KLASIFIKASI KEPUASAN SISWA TERHADAP KINERJA GURU MENGUNAKAN ALGORITMA SVM, C4.5, DAN RANDOM FOREST

Shona Chayy Bilqisth¹⁾, Rohmatulloh Muhamad Ikhsanuddin²⁾

¹⁾ Sistem Informasi, Fakultas Matematika dan IPA, Universitas Negeri Semarang,

²⁾ Sains Data, Fakultas Sains dan Teknologi, Universitas Putra Bangsa

¹⁾Kampus UNNES Sekaran, Kec. Gn. Pati, Kota Semarang, Jawa Tengah 50229,

²⁾Jl. Ronggowarsito No.18, Sudagaran, Kedawung, Kec. Pejagoan, Kabupaten Kebumen,
Jawa Tengah 54361

E-mail : ¹⁾shona@mail.unnes.ac.id , ²⁾ikhsanuddin@fst.universitaspuprabangsa.ac.id

ABSTRAK

Penelitian ini bertujuan membandingkan akurasi dan performa tiga algoritma klasifikasi, yaitu SVM (Support Vector Machine), C4.5, dan Random Forest, dalam mengklasifikasikan tingkat kepuasan siswa terhadap kinerja guru di MA Miftahussalam Demak. Data diperoleh dari kuesioner siswa dengan atribut seperti penguasaan materi, penjelasan materi, disiplin waktu, kepedulian, dan komunikasi. Preprocessing dilakukan dengan metode one-hot encoding dan normalisasi, serta evaluasi model menggunakan 10-fold cross-validation. Hasil menunjukkan bahwa SVM mencapai akurasi tertinggi sebesar 98,30%, diikuti oleh C4.5 sebesar 96,61%, dan Random Forest sebesar 95,76%. Uji ANOVA membuktikan perbedaan signifikan antar algoritma ($p < 0,05$), yang menandakan bahwa setidaknya satu algoritma memiliki performa yang berbeda secara statistik. Penelitian menyimpulkan bahwa SVM merupakan algoritma paling efektif untuk data ini, dengan rekomendasi penelitian lanjutan menguji algoritma ensemble dan dataset lebih besar.

Kata Kunci: SVM, C4.5, Random Forest, Klasifikasi, Kepuasan Siswa, Cross-validation.

ABSTRACT

This study aims to compare the accuracy and performance of three classification algorithms—Support Vector Machine (SVM), C4.5, and Random Forest—in classifying student satisfaction levels with teacher performance at MA Miftahussalam Demak. The data were obtained from student questionnaires featuring attributes such as subject mastery, clarity of explanation, punctuality, care, and communication. Preprocessing was carried out using one-hot encoding and normalization methods, and model evaluation was conducted using 10-fold cross-validation. The results showed that SVM achieved the highest accuracy at 98.30%, followed by C4.5 at 96.61%, and Random Forest at 95.76%. An ANOVA test confirmed a statistically significant difference among the algorithms ($p < 0.05$), indicating that at least one algorithm performs differently from the others. The study concludes that SVM is the most effective algorithm for this dataset, and recommends future research to explore ensemble algorithms and larger datasets.

Keywords: SVM, C4.5, Random Forest, Classification, Student Satisfaction, Cross-validation

1. PENDAHULUAN

1.1 Latar Belakang

Kepuasan siswa terhadap kinerja guru merupakan indikator utama dalam menilai

kualitas pendidikan yang diberikan. Evaluasi kinerja guru melalui umpan balik siswa memberikan wawasan penting bagi peningkatan kualitas pembelajaran. Mengukur tingkat kepuasan siswa bukan hanya sekadar untuk

mengetahui persepsi mereka, tetapi juga untuk mengidentifikasi faktor-faktor yang mempengaruhi proses belajar mengajar, seperti penguasaan materi, strategi pembelajaran, disiplin waktu, komunikasi, dan perhatian terhadap siswa. Oleh karena itu, pengukuran kepuasan siswa memiliki peran strategis dalam meningkatkan efektivitas pembelajaran di sekolah.

Seiring dengan perkembangan teknologi, metode yang dapat digunakan untuk menganalisis data kepuasan siswa adalah data mining, yang memungkinkan pengolahan dan analisis data dalam jumlah besar dan kompleks. Data mining dengan berbagai algoritma klasifikasi dapat membantu dalam memprediksi dan mengklasifikasikan tingkat kepuasan siswa berdasarkan data yang dikumpulkan, baik melalui kuesioner maupun umpan balik yang diberikan oleh siswa. Pada penelitian ini, peneliti membandingkan tiga algoritma yang umum digunakan dalam klasifikasi data, yaitu SVM (Support Vector Machine), C4.5, dan Random Forest.

Support Vector Machine (SVM) merupakan salah satu algoritma penting dalam *machine learning* yang digunakan untuk menyelesaikan masalah klasifikasi dan regresi. Klasifikasi pada SVM dapat dipandang sebagai sebuah masalah optimasi yang bertujuan menemukan model terbaik yang dapat merepresentasikan pola atau hubungan antara data dengan labelnya. SVM efektif dalam menangani data berdimensi tinggi dan non-linear dengan mencari hyperplane optimal untuk memisahkan kelas dalam ruang fitur. Kinerja SVM sangat dipengaruhi oleh parameter seperti C dan γ , yang berperan dalam mengatur toleransi kesalahan prediksi dan kompleksitas model pada SVM non-linear. SVM mampu memberikan akurasi tinggi dalam prediksi data kesehatan berdasarkan faktor risiko. Selain itu, teknik optimasi parameter seperti Grid Search dan Random Search sering digunakan untuk memperoleh nilai parameter yang optimal, dengan evaluasi performa dilakukan berdasarkan akurasi, penggunaan memori, dan waktu pengujian model terbaik pada berbagai dataset[1].

C4.5 adalah algoritma pohon keputusan yang membagi data berdasarkan atribut yang paling informatif. Pohon keputusan yang dihasilkan dari algoritma ini sangat mudah dipahami dan digunakan untuk menghasilkan aturan yang jelas dalam mengklasifikasikan data. C4.5 telah terbukti efektif dalam klasifikasi, seperti yang ditunjukkan oleh Tosun dan Bakan Kalaycioğlu (2024) dalam memprediksi prestasi akademik siswa di tingkat perguruan tinggi. Penelitian ini menyoroti keunggulan C4.5 dalam membangun pohon keputusan yang jelas dan mendalam, meskipun dalam aplikasi pendidikan terkait kepuasan siswa, algoritma ini memiliki keterbatasan dalam menghadapi data yang lebih kompleks dan besar dibandingkan dengan Random Forest[2].

Random Forest adalah algoritma ensemble yang menggabungkan banyak pohon keputusan untuk meningkatkan akurasi klasifikasi dan mengurangi risiko overfitting yang sering terjadi pada pohon keputusan tunggal. Algoritma ini sangat efektif dalam menangani dataset yang besar dan kompleks serta memberikan hasil yang lebih stabil dibandingkan algoritma pohon keputusan tunggal seperti C4.5. Berbagai penelitian menunjukkan bahwa Random Forest mampu memberikan performa yang kompetitif dalam berbagai aplikasi klasifikasi, termasuk dalam pengukuran kinerja dan evaluasi data yang kompleks[3].

Penelitian ini bertujuan untuk membandingkan efektivitas tiga algoritma klasifikasi, yaitu Support Vector Machine (SVM), C4.5, dan Random Forest, dalam mengukur tingkat kepuasan siswa terhadap kinerja guru di MA Miftahussalam Demak. Data yang digunakan dalam penelitian ini diperoleh dari kuesioner siswa yang mencakup berbagai atribut yang memengaruhi kepuasan, seperti penguasaan materi, penjelasan materi, disiplin waktu, kepedulian, dan komunikasi. Penelitian ini diharapkan dapat memberikan wawasan lebih mendalam mengenai efektivitas masing-masing algoritma dalam mengklasifikasikan kepuasan siswa, serta memberikan rekomendasi bagi pihak sekolah untuk meningkatkan kualitas pengajaran berdasarkan hasil analisis tersebut.

1.2 Tinjauan Pustaka

Penelitian mengenai penggunaan data mining untuk mengukur kepuasan siswa terhadap kinerja guru dan topik terkait lainnya telah banyak dilakukan. Dalam bidang klasifikasi, beberapa algoritma populer seperti Support Vector Machine (SVM), C4.5, dan Random Forest sering digunakan untuk mengolah data dan menghasilkan model prediksi yang akurat. Beberapa penelitian sebelumnya memberikan wawasan mengenai keunggulan dan kelemahan masing-masing algoritma dalam mengklasifikasikan berbagai jenis data.

Dewi (2023) menerapkan C4.5 untuk mengklasifikasikan kepuasan konsumen di Laundry Bunda. Dalam penelitian ini, C4.5 berhasil membangun pohon keputusan yang jelas dan informatif untuk mengukur kepuasan konsumen terhadap layanan yang diberikan. Hasil penelitian menunjukkan bahwa C4.5 dapat memberikan hasil yang stabil dalam menganalisis berbagai faktor yang mempengaruhi kepuasan konsumen, seperti harga dan kualitas layanan[4].

Mahendro (2022) meneliti kepuasan mahasiswa terhadap pembelajaran e-learning selama pandemi Covid-19 menggunakan algoritma Support Vector Machine (SVM). Data diperoleh dari kuesioner yang disebar kepada mahasiswa. Hasil analisis SVM menunjukkan akurasi sebesar 98,23%, dengan precision dan recall masing-masing 95,65%. Ini menunjukkan bahwa SVM mampu memprediksi responden yang puas dan tidak puas dengan tingkat ketepatan yang tinggi[5].

Penelitian oleh Pahlevi (2023) menunjukkan bahwa Random Forest mengungguli C4.5 dalam klasifikasi kelayakan kredit dengan akurasi yang lebih stabil dan memberikan hasil yang lebih optimal dibandingkan dengan algoritma lain seperti logistic regression dan support vector machine. Hasil penelitian menunjukkan bahwa Random Forest dapat memberikan nilai akurasi yang lebih baik dan kemampuan generalisasi yang lebih kuat[6].

Penelitian oleh Fitri (2024) menunjukkan bahwa Random Forest dan SVM kedua algoritma mampu melakukan klasifikasi dengan baik

setelah data disesuaikan. Penelitian ini memiliki tujuan untuk membandingkan performa dua algoritma klasifikasi, yakni Random Forest dan Support Vector Machine (SVM), dalam menganalisis sentimen terhadap opini masyarakat terkait pinjaman online di media sosial, khususnya di platform Twitter[7].

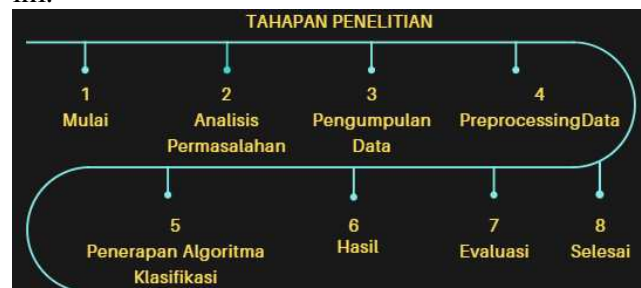
Tinjauan pustaka di atas memberikan gambaran yang jelas mengenai penerapan algoritma SVM, C4.5, dan Random Forest dalam berbagai konteks dan memberikan bukti empiris mengenai efektivitas masing-masing algoritma.

II. METODE PENELITIAN

Penelitian ini merupakan penelitian kuantitatif komparatif yang bertujuan untuk membandingkan efektivitas tiga algoritma klasifikasi, yaitu Support Vector Machine (SVM), C4.5, dan Random Forest dalam mengklasifikasikan kepuasan siswa terhadap kinerja guru.

2.1 Tahapan Penelitian

Proses penelitian ini dimulai dengan tahap analisis permasalahan, diikuti dengan pengumpulan data, dan diakhiri dengan penemuan hasil akhir. Gambar 1 menunjukkan tahapan-tahapan yang dilakukan dalam penelitian ini.



Gambar 1. Tahapan Penelitian

Proses penelitian dimulai dengan analisis permasalahan, yang bertujuan untuk mengidentifikasi isu utama dalam pengukuran kepuasan siswa. Tahap selanjutnya pengumpulan data dilakukan dengan menggunakan kuesioner yang diisi oleh siswa untuk mendapatkan informasi terkait atribut-atribut yang mempengaruhi kepuasan siswa terhadap kinerja guru. Data yang telah terkumpul kemudian diproses pada tahap *preprocessing* data, yang

meliputi pembersihan data dan normalisasi agar data siap digunakan dalam model klasifikasi.

Pada tahap penerapan algoritma klasifikasi SVM, C4.5, dan Random Forest, data yang telah diproses dianalisis menggunakan ketiga algoritma klasifikasi tersebut untuk mengukur tingkat kepuasan siswa. Hasil dari analisis ini akan menunjukkan perbandingan akurasi dari masing-masing algoritma, dengan tujuan untuk menentukan algoritma yang paling efektif dalam memprediksi kepuasan siswa. Tahap akhir dari penelitian ini adalah Hasil, di mana akurasi dari SVM, C4.5, dan Random Forest akan disajikan untuk menunjukkan performa masing-masing algoritma dalam klasifikasi kepuasan siswa. Penelitian ini diakhiri pada tahap Selesai, dengan kesimpulan yang memberikan rekomendasi berbasis hasil analisis untuk meningkatkan kualitas pengajaran di sekolah.

2.2 Data dan Teknik Pengumpulan Data

Data diperoleh melalui penyebaran kuesioner yang diisi oleh siswa MA Miftahussalam Demak. Kuesioner terdiri dari lima indikator utama: penguasaan materi, penjelasan materi, disiplin waktu, kepedulian, dan komunikasi. Respon siswa terhadap masing-masing indikator dikategorikan menjadi label “PUAS” dan “KURANG PUAS”.

2.3 Preprocessing Data

Proses *preprocessing* dilakukan untuk memastikan bahwa data dalam kondisi optimal sebelum digunakan oleh algoritma klasifikasi. Tahapan *preprocessing* yang dilakukan pada penelitian ini meliputi :

1. One-Hot Encoding, seluruh atribut kategorikal dalam dataset diubah menjadi bentuk numerik biner menggunakan filter NominalToBinary yang tersedia pada Weka 3.8.6. Teknik ini bertujuan agar algoritma seperti SVM dapat memproses data tanpa kendala, karena algoritma tersebut memerlukan input dalam bentuk numerik.
2. Normalisasi, data dinormalisasi ke dalam rentang nilai antara 0 hingga 1 dengan menggunakan filter Normalize pada Weka.

Langkah ini sangat penting, terutama untuk algoritma seperti SVM, yang performanya sangat dipengaruhi oleh perbedaan skala antar fitur. Normalisasi dilakukan menggunakan rumus sebagai berikut:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

di mana:

- x = nilai asli
 - x_{\min} = nilai minimum dalam kolom atribut
 - x_{\max} = nilai maksimum dalam kolom atribut
 - x' = nilai setelah normalisasi
3. Penetapan atribut target, atribut kelas atau label target ditetapkan pada kolom terakhir dalam dataset. Label ini terdiri dari dua kelas utama: “PUAS” dan “KURANG PUAS”, yang merepresentasikan tingkat kepuasan siswa terhadap kinerja guru. Proses klasifikasi bertujuan memprediksi label ini berdasarkan lima atribut penilaian utama yang diberikan siswa.

2.4 Pembagian Data

Penelitian ini menggunakan metode 10-fold cross-validation untuk menguji performa model. Dalam metode ini, dataset dibagi menjadi 10 bagian yang seimbang. Model dilatih pada 9 bagian dan diuji pada bagian ke-10 secara bergiliran hingga seluruh data digunakan sebagai data uji. Cross-validation dipilih agar evaluasi model lebih stabil dan tidak tergantung pada satu kali pembagian data.

2.4 Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah algoritma pembelajaran mesin yang digunakan untuk tugas klasifikasi dan regresi. SVM bekerja dengan mencari hyperplane optimal yang dapat memisahkan dua kelas data dengan margin maksimum, yaitu jarak terbesar antara hyperplane dan data dari masing-masing kelas yang terdekat, yang disebut sebagai support vectors. Secara umum, tujuan utama dari SVM adalah memaksimalkan margin antara dua kelas dengan meminimalkan kesalahan klasifikasi.

Dalam bentuk matematis, SVM menyelesaikan masalah optimasi berikut:

$$\min_{w,b} \frac{1}{2} \|w\|^2$$

di mana :

w adalah vektor bobot

b adalah bias

2.5 Algoritma C.45

C4.5 adalah algoritma pohon keputusan yang membangun pohon keputusan berdasarkan atribut yang paling informatif menggunakan gain informasi. Algoritma ini menghasilkan pohon keputusan yang mudah dipahami.

Rumus:

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

2.6 Random Forest

Random Forest adalah algoritma ensemble yang menggabungkan banyak pohon keputusan untuk meningkatkan akurasi klasifikasi dan mengurangi overfitting. Hasil klasifikasi diperoleh dari mayoritas suara pohon-pohon keputusan.

Rumus:

$$\hat{y} = \frac{1}{N} \sum_{i=1}^N f_i(X)$$

2.7 Evaluasi Model

Evaluasi dilakukan menggunakan metrik-metrik klasifikasi sebagai berikut:

1. Akurasi, proporsi data yang diklasifikasikan dengan benar:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

2. Precision, ketepatan klasifikasi positif :

$$\text{Precision} = \frac{TP}{TP + FP}$$

3. Recall, kemampuan model menemukan semua instance positif :

$$\text{Recall} = \frac{TP}{TP + FN}$$

4. F1-Score, Rata-rata harmonik antara precision dan recall:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

5. Kappa Statistic, mengukur sejauh mana klasifikasi lebih baik dibandingkan tebakan acak. Nilai kappa antara 0,81–1 menunjukkan kekuatan klasifikasi yang sangat baik.

2.7 Uji Anova

Uji Statistik ANOVA Untuk melengkapi analisis performa klasifikasi, dilakukan uji statistik menggunakan One-Way ANOVA (Analysis of Variance). Uji ini bertujuan untuk mengidentifikasi apakah terdapat perbedaan yang signifikan secara statistik antara rata-rata akurasi dari ketiga algoritma klasifikasi (SVM, C4.5, dan Random Forest) yang diuji menggunakan 10-fold cross-validation. Hasil akurasi dari masing-masing fold digunakan sebagai input dalam uji ANOVA. Perhitungan dilakukan menggunakan Python dengan bantuan pustaka SciPy. Hipotesis nol (H0) menyatakan bahwa tidak terdapat perbedaan signifikan antara ketiga algoritma, sementara hipotesis alternatif (H1) menyatakan bahwa terdapat setidaknya satu perbedaan signifikan.

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Data yang digunakan dalam studi ini diperoleh dari kuesioner yang disebarikan kepada siswa Madrasah Aliyah Miftahussalam Demak. Penulis mengambil sampel sebanyak 118 responden yang berasal dari tiga kelas siswa MA Miftahussalam untuk dianalisis.

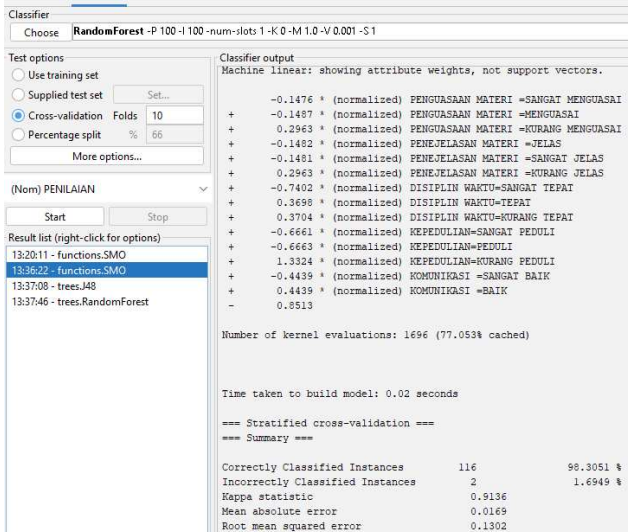
Tabel 1 Data Penelitian

Nama	Kelas	K1	K2	K3	K4	K5	P/TP
ssw1	x	Sangat Menguasai	Jelas	Sangat Tepat	Sangat Peduli	Sangat Baik	Puas
ssw2	X	Menguasai	Jelas	Tepat	Sangat Peduli	Baik	Puas
Ssw3	X	Sangat Menguasai	Sangat Jelas	Sangat Tepat	Sangat Peduli	Sangat Baik	Puas
ssw 118	X	Menguasai	Kurang Jelas	Tepat	Kurang Peduli	Baik	Kurang Puas

Data yang telah dikumpulkan selanjutnya dikonversi ke dalam format .csv untuk memudahkan implementasi dan analisis menggunakan aplikasi Weka 3.8.6.

3.2 Penerapan Algoritma Klasifikasi dengan SVM, C.45, dan Random Forest

Pada studi ini, tiga algoritma klasifikasi, yaitu SVM, C4.5, dan Random Forest, digunakan untuk mengukur tingkat kepuasan siswa terhadap kinerja guru di Madrasah Aliyah Miftahussalam Demak. Data yang digunakan dalam studi ini diperoleh dari kuesioner yang diisi oleh 118 responden, yang berasal dari tiga kelas siswa. Setiap algoritma diuji untuk mengukur Precision, Recall, Akurasi, dan Kappa Statistic. Proses data dan implementasi model klasifikasi dilakukan menggunakan Weka 3.8.6, sebuah perangkat lunak yang digunakan untuk analisis data dan penerapan algoritma machine learning. Ditunjukkan pada gambar 2 berikut ini :



Gambar 2. Tampilan Pada Aplikasi WEKA

3.3 Hasil Klasifikasi

Penelitian ini bertujuan untuk membandingkan efektivitas tiga algoritma klasifikasi, yaitu Support Vector Machine (SVM), C4.5 (J48), dan Random Forest dalam mengklasifikasikan tingkat kepuasan siswa terhadap kinerja guru. Klasifikasi dilakukan terhadap data hasil kuesioner yang memuat lima atribut utama: penguasaan materi, penjelasan materi, disiplin waktu, kepedulian, dan komunikasi. Proses klasifikasi dilakukan melalui perangkat lunak Weka 3.8.6 dengan pendekatan evaluasi 10-fold cross-validation, serta preprocessing berupa one-hot encoding dan normalisasi terhadap atribut kategorikal.

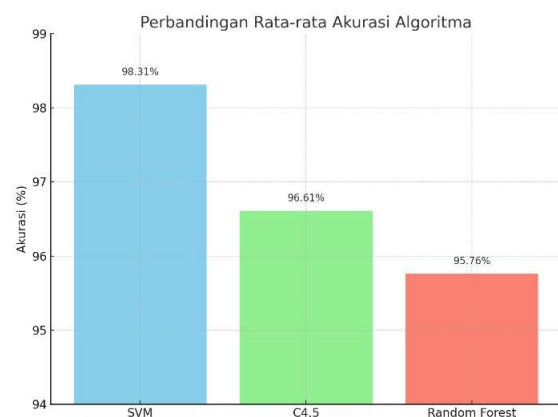
Setelah semua tahapan *preprocessing* dan pengaturan parameter algoritma dilakukan, ketiga model klasifikasi dijalankan dan menghasilkan output performa masing-masing berdasarkan lima indikator evaluasi: akurasi, precision, recall, F1-score, dan kappa statistic.

Hasil pengujian terhadap masing-masing algoritma dapat dilihat pada tabel 2 berikut :

Tabel 2 Perbandingan Hasil Akurasi, Precision, dan Recall

Algoritma	Akurasi	Kappa	Precision	Recall	F1-Score
SVM (SMO)	98.31%	0.9136	0.983	0.983	0.982
C4.5 (J48)	96.61%	0.8379	0.966	0.966	0.966
Random Forest	95.76%	0.7909	0.957	0.958	0.957

Dari Tabel 2, dapat dilihat bahwa SVM memperoleh akurasi tertinggi sebesar 98.31% dan kappa statistic sebesar 0.9136, yang mengindikasikan bahwa model memiliki klasifikasi yang sangat kuat dibandingkan dengan tebakan acak. Algoritma ini juga menunjukkan nilai precision dan recall seimbang (0.983), menjadikannya algoritma dengan performa paling konsisten dalam eksperimen ini. Berikut ini gambar perbandingan rata-rata akurasi algoritma.



Gambar 3. Perbandingan Rata-rata Akurasi Algoritma

3.4 Evaluasi

Untuk melengkapi analisis hasil klasifikasi, dilakukan evaluasi berdasarkan confusion matrix dari masing-masing algoritma.

Confusion matrix memberikan gambaran rinci mengenai jumlah prediksi yang benar dan salah pada masing-masing kelas, sehingga memungkinkan penelusuran lebih dalam terhadap performa model, khususnya dalam membedakan antara dua kelas: “PUAS” dan “KURANG PUAS”.

Tabel 3 berikut merangkum hasil evaluasi confusion matrix dari ketiga algoritma yang diuji:

Tabel 3 confusion matrix

Algoritma	Kelas	True	False
SVM (SMO)	PUAS	104	0
	KURANG PUAS	12	2
C4.5 (J48)	PUAS	102	2
	KURANG PUAS	12	2
Random Forest	PUAS	102	2
	KURANG PUAS	11	3

1. SVM (SMO) menunjukkan performa terbaik secara keseluruhan:
 - a. 100% akurat dalam memprediksi kelas PUAS (104 dari 104)
 - b. Hanya 2 kesalahan klasifikasi pada kelas KURANG PUAS
 - c. Ini sejalan dengan metrik evaluasi lainnya seperti F1-score (0.982) dan kappa statistic (0.9136)
2. C4.5 (J48) memiliki jumlah kesalahan moderat, yaitu total 4 instance:
 - a. Dua kesalahan pada masing-masing kelas
 - b. Performa tetap tinggi dengan F1-score sebesar 0.966 dan kappa statistic 0.8379
 - c. Keunggulan C4.5 terletak pada interpretabilitas aturan klasifikasinya
3. Random Forest mencatat kesalahan terbanyak di antara ketiganya, yaitu 5 instance:
 - a. Dua kesalahan klasifikasi pada kelas PUAS
 - b. Tiga kesalahan pada kelas KURANG PUAS
 - c. Kappa statistic-nya yang paling rendah (0.7909) mengindikasikan bahwa model ini meskipun cukup akurat, kurang stabil dibanding dua model lainnya dalam kasus ini.

Analisis berdasarkan confusion matrix memperkuat kesimpulan bahwa SVM adalah

algoritma paling efektif untuk dataset ini, dengan jumlah klasifikasi benar tertinggi dan kesalahan paling minim. C4.5 menunjukkan kinerja yang baik dan seimbang, menjadikannya alternatif yang dapat dipertimbangkan terutama ketika interpretasi model penting. Random Forest, meskipun secara teoritis unggul dalam data kompleks dan besar, dalam penelitian ini menghasilkan performa sedikit lebih rendah, kemungkinan disebabkan oleh ukuran dataset yang relatif kecil.

3.5 Hasil Uji Statistik Anova

Hasil uji Anova menunjukkan bahwa nilai $F = 821.58$ dan $p\text{-value} = 6.55 \times 10^{-25}$. Nilai p yang lebih kecil dari 0.05 ini menandakan bahwa terdapat perbedaan yang signifikan secara statistik antara akurasi rata-rata ketiga algoritma. Dengan demikian, hipotesis nol ditolak, dan disimpulkan bahwa performa ketiga algoritma tidak setara secara statistik. Hasil ini memperkuat temuan sebelumnya bahwa SVM memberikan performa klasifikasi terbaik dibandingkan dua algoritma lainnya dalam konteks *dataset* ini.

4. KESIMPULAN DAN SARAN

4.1 Kesimpulan

Penelitian ini bertujuan untuk membandingkan efektivitas tiga algoritma klasifikasi, yaitu Support Vector Machine (SVM), C4.5, dan Random Forest, dalam mengklasifikasikan tingkat kepuasan siswa terhadap kinerja guru di MA Miftahussalam Demak berdasarkan atribut penguasaan materi, penjelasan materi, disiplin waktu, kepedulian, dan komunikasi. Evaluasi dilakukan menggunakan metode 10-fold cross-validation di Weka 3.8.6 dengan preprocessing berupa one-hot encoding dan normalisasi.

Hasil analisis menunjukkan bahwa:

1. SVM (SMO) memberikan hasil terbaik dengan akurasi sebesar 98.31%, precision dan recall seimbang sebesar 0.983, serta nilai Kappa Statistic 0.9136, yang menunjukkan kekuatan klasifikasi sangat tinggi.
2. C4.5 (J48) juga menunjukkan performa yang baik dengan akurasi 96.61%, serta kemampuan interpretasi model yang unggul melalui pohon keputusan yang jelas.

3. Random Forest, meskipun secara umum dikenal kuat, pada dataset ini memperoleh akurasi 95.76%, sedikit di bawah dua algoritma lainnya, dengan kappa statistic 0.7909.

Dengan demikian, dapat disimpulkan bahwa SVM adalah algoritma paling efektif untuk klasifikasi kepuasan siswa pada konteks dan data ini, baik dari segi akurasi maupun stabilitas prediksi.

4.2 Saran

1. Implementasi SVM, algoritma SVM dapat digunakan sebagai dasar pengembangan sistem evaluasi kinerja guru berbasis data yang akurat dan efisien.
2. Perluasan *dataset*, penelitian lanjutan disarankan menggunakan dataset yang lebih besar dan variatif untuk meningkatkan generalisasi model.
3. Eksplorasi Algoritma Lain, peneliti berikutnya dapat membandingkan SVM dengan algoritma lain seperti XGBoost atau voting ensemble untuk memperoleh hasil yang lebih optimal.
4. Hasil uji Anova mendukung hasil evaluasi individual masing-masing algoritma. Untuk mengetahui lebih lanjut algoritma mana yang berbeda secara signifikan, uji post-hoc seperti Tukey HSD dapat dipertimbangkan dalam penelitian lanjutan.
5. Pemanfaatan Hasil dalam Kebijakan Sekolah, hasil klasifikasi dapat dimanfaatkan sebagai bahan pertimbangan dalam pengambilan keputusan peningkatan mutu pembelajaran.

5. DAFTAR PUSTAKA

- [1] M. Fajri and A. Primajaya, “Komparasi Teknik Hyperparameter Optimization pada SVM untuk Permasalahan Klasifikasi dengan Menggunakan Grid Search dan Random Search,” *J. Appl. Informatics Comput.*, vol. 7, no. 1, pp. 14–19, Jan. 2023.
- [2] S. Tosun and D. Bakan Kalaycıoğlu, “Data mining approach for prediction of academic success in open and distance education,” *J. Educ. Technol. Online*

- Learn.*, vol. 7, no. 2, pp. 168–176, 2024.
- [3] D. E. Asuquo, U. A. Umoh, F. B. Osang, and E. W. Okokon, “Performance Evaluation of C4.5, Random Forest and Naïve Bayes Classifiers in Employee Performance and Promotion Prediction,” © *Afr. J. MIS*, vol. 2, no. 4, pp. 41–55, 2020.
- [4] R. Dewi, Z. Hanif, and I. Santoso, “Analisa Kepuasan Konsumen Pada Laundry Menggunakan Algoritma C4. 5,” *IKRA-ITH Inform. J. Komput. dan Inform.*, vol. 7, no. 2, pp. 133–141, 2023.
- [5] I. Mahendro and D. Abimanto, “Analisa Kepuasan Mahasiswa Terhadap E-Learning Menggunakan Algoritma Support Vector Machine,” *J. SAINS DAN Teknol. Marit.*, vol. 23, no. 1, p. 97, Sep. 2022.
- [6] O.- Pahlevi, A.- Amrin, and Y.- Handrianto, “Implementasi Algoritma Klasifikasi Random Forest Untuk Penilaian Kelayakan Kredit,” *J. Infotech*, vol. 5, no. 1, pp. 71–76, 2023.
- [7] D. A. Fitri, “Komparasi algoritma random forest classifier dan support vector machine untuk sentimen masyarakat terhadap pinjaman online di media sosial,” vol. 9, no. 4, pp. 2018–2029, 2024.