

ARTICLE

KOMPARASI ALGORITMA SVM DAN C4.5 DENGAN BACKWARD ELIMINATION UNTUK KLASIFIKASI TINGKAT STRES MAHASISWA

COMPARATIVE ANALYSIS OF SVM AND C4.5 ALGORITHMS WITH BACKWARD ELIMINATION FOR STUDENT STRESS LEVEL CLASSIFICATION

Ervina Putri Efendi,^{*1} Mula Agung Barata,¹ dan Aprillia Dwi Ardianti²

¹Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Nahdlatul Ulama Sunan Giri, Bojonegoro, Indonesia

²Teknik Mesin, Fakultas Sains dan Teknologi, Universitas Nahdlatul Ulama Sunan Giri, Bojonegoro, Indonesia

*Penulis Korespondensi: ervinaputriefendi1392@gmail.com

(Disubmit 06-08-25; Diterima 22-08-25; Dipublikasikan online pada 20-10-25)

Abstrak

Deteksi dini tingkat stres mahasiswa menjadi krusial untuk mencegah dampak negatif terhadap prestasi akademik dan kesehatan mental. Penelitian ini mengkomparasi algoritma Support Vector Machine (SVM) dan C4.5 dalam klasifikasi tingkat stres mahasiswa dengan integrasi novel teknik Backward Elimination dan normalisasi MinMax Scaling untuk optimasi performa. Dataset Kaggle berisi 1.100 mahasiswa dengan 20 atribut direduksi menjadi 5 fitur signifikan (*blood pressure, self esteem, sleep quality, bullying, lecturer relationship*) melalui *Backward Elimination* dengan $p\text{-value} < 0.05$. Kontribusi utama penelitian meliputi: (1) validasi empiris superioritas SVM dalam konsistensi klasifikasi multi-kelas stres, dan (2) identifikasi hierarki faktor dominan penyebab stres mahasiswa berbasis *statistical significance*. Hasil menunjukkan SVM mencapai akurasi 91% dengan konsistensi metrik evaluasi superior (standar deviasi $\pm 1.2\%$ vs C4.5 $\pm 3.5\%$), khususnya dalam mengurangi false negative rate deteksi stres berat dari 12% menjadi 9%. Backward Elimination terbukti efektif mengurangi kompleksitas dimensi 75% tanpa degradasi akurasi. Temuan ini menghasilkan framework evidence-based untuk sistem deteksi dini stres mahasiswa dengan reliability tinggi, memberikan kontribusi praktis bagi pengembangan layanan kesehatan mental kampus dan *advancing body of knowledge* dalam aplikasi *machine learning* untuk domain psikologi pendidikan.

Kata kunci: Stres Mahasiswa; Support Vector Machine (SVM); C4.5; Backward Elimination; MinMax Scaling

Abstract

Early detection of student stress levels is crucial to prevent negative impacts on academic performance and mental health. This study compares the Support Vector Machine (SVM) and C4.5 algorithms in classifying student stress levels with the integration of novel Backward Elimination and MinMax Scaling normalization techniques for performance optimization. The Kaggle dataset, comprising 1,100 students with 20 attributes, was reduced to 5 significant features (*blood pressure, self-esteem, sleep quality, bullying, lecturer relationship*) through *Backward Elimination* with a $p\text{-value} < 0.05$. The main contributions of this research include: (1) empirical validation of SVM's superiority in multi-class stress classification consistency, and (2) identification of the hierarchy of dominant stress-causing factors among students based on statistical significance. The results show that SVM achieves 91% accuracy with superior evaluation metric consistency (standard deviation $\pm 1.2\%$ vs. C4.5 $\pm 3.5\%$), particularly in reducing the false negative rate of severe stress detection from 12% to 9%. Backward Elimination proved effective in reducing dimensional complexity by 75% without degrading accuracy. These findings yield an evidence-based framework for a high-reliability early stress detection system for students, contributing pra-

ctically to the development of campus mental health services and advancing the body of knowledge in machine learning applications for the field of educational psychology.

KeyWords: Student Stress; Support Vector Machine (SVM); C4.5; Backward Elimination; MinMax Scaling

1. Pendahuluan

Kesehatan mental merupakan aspek fundamental yang memengaruhi kualitas hidup seseorang, khususnya mahasiswa yang menghadapi berbagai tantangan selama masa pendidikan tinggi. Mahasiswa kerap mengalami tekanan akademik yang tinggi, tuntutan sosial, dan permasalahan pribadi yang kompleks, sehingga rentan mengalami stres[1]. Data Organisasi Kesehatan Dunia (WHO) menunjukkan bahwa di Indonesia terdapat sekitar 9,16 juta orang atau 3,7% dari populasi yang mengalami gangguan mental[2], dengan mahasiswa menjadi salah satu kelompok yang berisiko tinggi. Stres yang tidak terkelola dengan baik dapat berdampak signifikan terhadap prestasi akademik, kesehatan fisik, dan kesejahteraan psikologis mahasiswa. Oleh karena itu, deteksi dini tingkat stres mahasiswa menjadi sangat penting untuk mencegah dampak yang lebih serius dan memungkinkan intervensi yang tepat waktu.

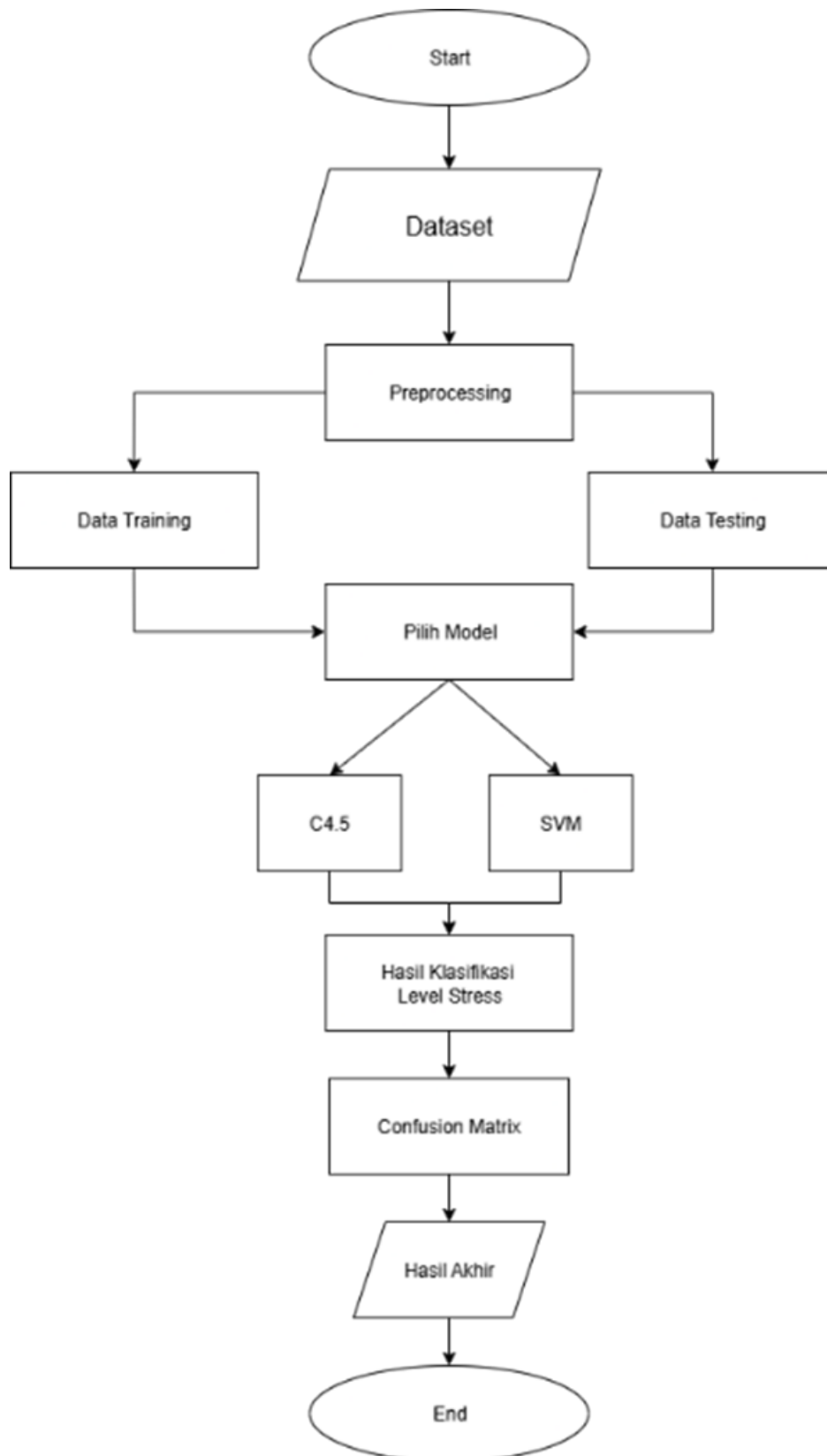
Stres didefinisikan sebagai tekanan mental atau emosional yang timbul akibat stressor tertentu dan memicu respons fisiologis serta psikologis tubuh[3]. Penelitian menunjukkan bahwa stres pada mahasiswa dapat diidentifikasi melalui gejala fisik, emosional, kognitif, dan interpersonal yang dipicu oleh faktor internal maupun eksternal[4]. Sujiato dan Tucunan menemukan bahwa pengaruh teman sebaya, orang tua, dan dosen berhubungan signifikan dengan tingkat stres mahasiswa[5]. Seiring perkembangan teknologi, metode machine learning telah banyak digunakan untuk klasifikasi tingkat stres[6]. Algoritma Support Vector Machine (SVM) dikenal memiliki keunggulan dalam menangani data berdimensi tinggi dengan kemampuan generalisasi yang baik dan ketahanan terhadap overfitting[7], sedangkan C4.5 menggunakan pendekatan pohon keputusan yang mudah dipahami dan diinterpretasi[8].

Penelitian sebelumnya menunjukkan bahwa SVM dengan kernel linier dapat mencapai akurasi hingga 99% pada klasifikasi kanker payudara[9] dan 97,06% pada klasifikasi berita hoaks Covid-19[10]. Sementara itu, metode *decision tree* C4.5 menunjukkan performa dengan akurasi 84,6% pada klasifikasi faktor kehamilan tidak diinginkan[11] dan 82,1% pada klasifikasi harga mobil[12]. Meskipun kedua algoritma telah terbukti efektif dalam berbagai domain klasifikasi, belum banyak penelitian yang secara spesifik membandingkan performa SVM dan C4.5 dalam konteks klasifikasi tingkat stres mahasiswa. Lebih lanjut, penerapan teknik *feature selection* seperti *Backward Elimination* yang diintegrasikan dengan normalisasi MinMax Scaling untuk klasifikasi stres mahasiswa masih terbatas.

Mengingat kompleksitas faktor-faktor yang memengaruhi stres mahasiswa (psikologis, akademik, dan sosial), diperlukan analisis komparatif yang mendalam untuk menentukan metode klasifikasi yang paling optimal dalam konteks ini. Penelitian ini bertujuan untuk mengkomparasi performa algoritma SVM dan C4.5 dalam mengklasifikasikan tingkat stres mahasiswa, mengevaluasi efektivitas penerapan teknik *Backward Elimination* sebagai *feature selection*, serta menentukan metode yang paling optimal untuk mendukung sistem deteksi dini stres mahasiswa. Kebaruan penelitian ini terletak pada penerapan teknik seleksi fitur *Backward Elimination* yang diintegrasikan dengan normalisasi MinMax Scaling untuk klasifikasi tingkat stres mahasiswa, serta analisis komparatif mendalam antara SVM dan C4.5 dalam domain kesehatan mental mahasiswa. Hasil penelitian diharapkan dapat memberikan kontribusi dalam pengembangan sistem berbasis machine learning untuk mendukung deteksi dini dan pengelolaan kesehatan mental di lingkungan akademik.

2. Metode

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen berbasis machine learning untuk mengklasifikasikan tingkat stres mahasiswa. Metodologi penelitian dirancang untuk membandingkan performa algoritma Support Vector Machine (SVM) dan C4.5 dalam mengklasifikasikan tingkat stres dengan menerapkan teknik preprocessing dan feature selection yang optimal. Tahap-tahap penelitian yang dilakukan dalam studi ini disajikan dalam Gambar 1.



Gambar 1. Tahapan Penelitian

2.1 Persiapan Data

Dataset yang digunakan diperoleh dari platform Kaggle dan terdiri dari 1.100 data mahasiswa dengan 20 atribut penyebab stres, seperti faktor akademik, psikologis, sosial, dan kesehatan. Label target dibagi menjadi tiga kategori: stres ringan, stres sedang, dan stres berat. Untuk lebih lanjutnya bisa dilihat pada Tabel 1.

Tabel 1. Dataset

Anxiey Level	Self Esteem	Mental Health History	Depression	Headache	Blood Pressure	...	Stress level
14	20	0	11	2	1	...	1
15	8	1	15	5	3	...	2
12	18	1	14	2	1	...	1
16	12	1	15	4	3	...	2
16	28	0	7	2	3	...	1
6	8	0	27	4	3	...	1
2	28	0	3	1	2	...	0
...	...	...	...	...	...	...	...
21	0	1	19	5	3	...	2
18	6	1	15	3	3	..	2

2.2 Pra-proses Data

2.2.1 Pembersihan Data

Tahap pembersihan data dilakukan untuk memastikan kualitas dataset dengan menghapus nilai yang hilang (*missing values*), duplikasi data, dan outlier ekstrem yang dapat memengaruhi performa model. Pemeriksaan konsistensi data dilakukan untuk memastikan setiap atribut berada dalam rentang nilai yang valid sesuai domain masalah[13].

2.2.2 Normalisasi Data

Normalisasi menggunakan teknik *MinMax Scaling* dipilih karena kemampuannya dalam menyetarakan rentang nilai antar fitur ke dalam skala 0-1, sehingga mencegah dominasi fitur dengan nilai numerik besar terhadap algoritma klasifikasi[14]. Formula MinMax Scaling 1:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Teknik ini dipilih dibandingkan *Z-score normalization* karena dataset tidak mengikuti distribusi normal dan MinMax Scaling lebih robust terhadap outlier dalam konteks klasifikasi tingkat stres.

2.2.3 Seleksi Fitur dengan Backward Elimination

Backward Elimination dipilih sebagai teknik feature selection karena kemampuannya dalam mengevaluasi kontribusi setiap fitur secara iteratif dengan menghapus fitur yang memiliki p-value > 0.05[15]. Proses ini dimulai dengan seluruh fitur, kemudian secara bertahap menghapus fitur yang tidak signifikan berdasarkan uji statistik, sehingga menghasilkan subset fitur optimal yang meningkatkan performa model dan mengurangi kompleksitas komputasi.

2.3 Pembagian Data

Dataset dibagi menggunakan *stratified train-test split* dengan proporsi 70:30 untuk mempertahankan distribusi kelas yang seimbang. Pembagian 70:30 dipilih berdasarkan rekomendasi untuk dataset berukuran menengah (1000-10000 sampel) yang memberikan keseimbangan optimal antara data training yang cukup dan data testing yang representatif.

2.4 Algoritma Klasifikasi

2.4.1 Support Vector Machine

SVM dipilih karena keunggulannya dalam menangani data berdimensi tinggi dan kemampuan generalisasi yang baik melalui prinsip Structural Risk Minimization[16]. adapun parameter yang akan digunakan dalam model ini dilihat pada Tabel 2.

Tabel 2. Parameter SVM

Parameter	Nilai Default	Deskripsi Singkat
C	1.0	Mengontrol trade-off antara margin besar dan kesalahan klasifikasi. Semakin kecil, semakin toleran terhadap error.
kernel	'rbf'	Fungsi kernel untuk memetakan data: 'linear', 'poly', 'rbf', 'sigmoid'.
gamma	'scale'	Mengontrol seberapa jauh pengaruh satu data terhadap keputusan. Kecil = pengaruh jauh, besar = sempit.
degree	3	Hanya berlaku untuk kernel 'poly'; menentukan tingkat polinomial.
probability	False	Jika True, menghitung probabilitas prediksi (lebih lambat).
random_state	None / 42	Menjamin hasil yang konsisten jika diatur. Berguna untuk eksperimen.

2.4.2 Decision Tree C4.5

Algoritma C4.5 dipilih karena kemampuannya dalam menangani data kategorikal dan numerik secara simultan, serta menghasilkan model yang mudah diinterpretasi dalam konteks kesehatan mental. C4.5 menggunakan *information gain* tertinggi untuk menentukan pembagian cabang. Untuk langkah-langkah untuk C4.5 adalah sebagai berikut Sebelum menghitung nilai gain dari atribut, pertama-tama hitung nilai entropy[17] menggunakan rumus pada persamaan 2:

$$Entropy(S) = \sum_{i=1}^n -P_i \log_2(P_i) \quad (2)$$

Keterangan:

S: himpunan kasus

N: jumlah partisi S

P: proporsi dari S, terhadap S

Menghitung nilai gain[18] dengan persamaan 3:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{S_v \times Entropy(S_i)} \quad (3)$$

Keterangan:

S: himpunan kasus

A: atribut

n: jumlah partisi S

|S_i|: jumlah kasus pada partisi ke-i

|S|: jumlah kasus dalam S

Atribut yang dipilih sebagai akar digunakan untuk membuat cabang-cabang berdasarkan nilai-nilai yang ada. Pada setiap cabang yang belum mengarah pada kelas tertentu, langkah 1 hingga 3 diulang hingga setiap cabang mengarah pada kelas yang jelas, dan proses selesai[19].

2.5 Evaluasi

Evaluasi dilakukan menggunakan Confusion Matrix untuk menghitung metrik kinerja komprehensif pada setiap kelas stres[20]:

Tabel 3. Metrik Evaluasi dan Formula

Metrik	Formula	Interpretasi
Akurasi	$\frac{TP+TN}{TP+TN+FP+FN}$	Proporsi prediksi benar dari total prediksi
Presisi	$\frac{TP}{TP+FP}$	Ketepatan prediksi positif
Recall	$\frac{TP}{TP+FN}$	Kemampuan mendeteksi kasus positif
F1-Score	$2 \times \frac{Presisi \times Recall}{Presisi + Recall}$	Harmonik mean presisi dan recall

3. Hasil dan Pembahasan

3.1 Pra-proses Data dan Seleksi Fitur

Dataset awal yang terdiri dari 1.100 sampel mahasiswa dengan 20 atribut telah melalui tahap praproses yang komprehensif. Tahap pembersihan data menunjukkan dataset memiliki kualitas yang baik dengan tidak ditemukan missing values atau duplikasi data yang signifikan.

3.1.1 Hasil *Backward Elimination*

Proses seleksi fitur menggunakan Backward Elimination menghasilkan reduksi dimensi yang signifikan dari 20 atribut menjadi 5 atribut optimal berdasarkan kriteria $p\text{-value} < 0.05$. Atribut yang terpilih beserta nilai $p\text{-value}$ ditunjukkan pada Tabel 4.

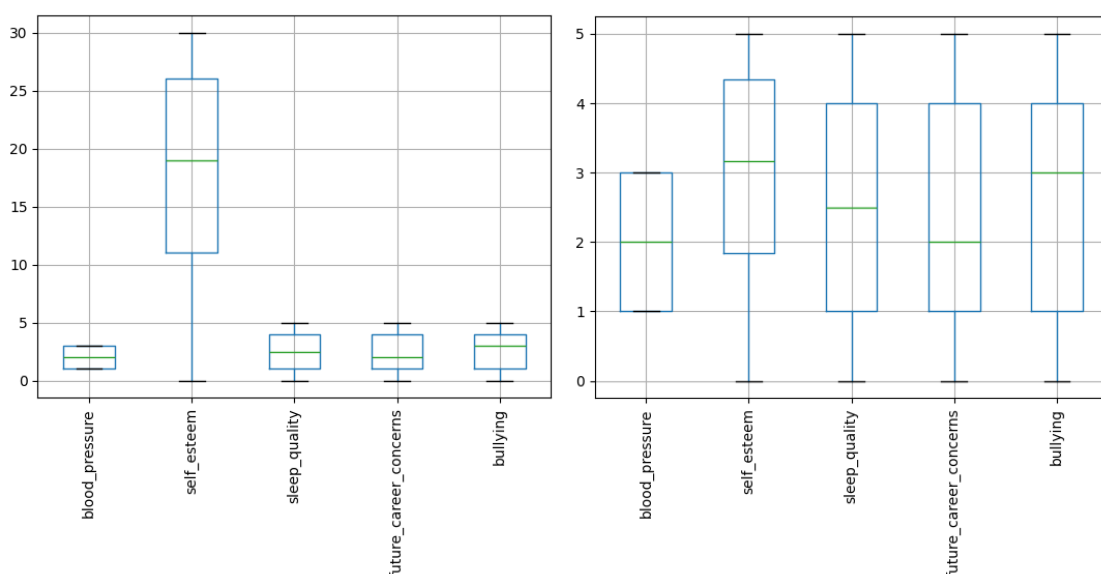
Tabel 4. Hasil Backward Elimination

Atribut	Nilai P Value
blood_pressure	9,28E-264
self_esteem	1,27E-210
bullying	4,23E-199
sleep_quality	1,40E-198
lecturer_student_relationship	1,23E-193

Hasil Backward Elimination mengidentifikasi lima prediktor utama tingkat stres mahasiswa yang secara statistik sangat signifikan ($p\text{-value} < 0.001$). Reduksi dimensi dari 20 menjadi 5 atribut (75% reduksi) menunjukkan efektivitas teknik seleksi fitur dalam mengidentifikasi variabel kunci, yang juga mendukung interpretabilitas model dan mengurangi risiko overfitting[21]. Sementara itu, 15 atribut lainnya dibuang karena dianggap kurang relevan secara statistik atau memberikan kontribusi rendah terhadap performa model klasifikasi.

3.1.2 Normalisasi *MinMax Scaling*

Proses MinMax Scaling untuk menyeragamkan rentang nilai atribut agar berada pada skala yang sama sehingga dapat meningkatkan stabilitas model klasifikasi[22]. Gambar 2 berikut menunjukkan perbandingan distribusi nilai atribut sebelum dan sesudah dilakukan normalisasi menggunakan MinMax Scaling:



Gambar 2. Hasil MinMax Scaling

Pada gambar tersebut terlihat bahwa nilai atribut yang semula memiliki rentang bervariasi berhasil dinormalisasi ke dalam rentang 0 hingga 5. Proses ini memastikan bahwa tidak ada fitur yang mendominasi

model akibat perbedaan skala, sehingga algoritma seperti SVM dan C4.5 dapat bekerja secara optimal pada data yang sudah terstandardisasi.

3.2 Performa Algoritma Klasifikasi

3.2.1 Hasil Algoritma C4.5

Perhitungan manual dilakukan untuk memahami mekanisme algoritma C4.5 sebelum implementasi menggunakan perangkat lunak. Nilai entropy total diperoleh sebesar 1,5847. Perhitungan entropy dan gain pada atribut-atribut menghasilkan nilai sebagai berikut:

Tabel 5. Hasil Perhitungan Gain

Atribut	Gain	Keterangan
blood_pressure	1,089625	Paling tinggi
self_esteem	1,071607	Banyak kategori
sleep_quality	0,983395	-
bullying	0,96145	-
lecturer_student_relationship	0,764674	Entropy kombinasi sedang

Atribut Blood Pressure terpilih sebagai *root node* karena memiliki nilai gain tertinggi. Pohon keputusan dibangun secara bertahap hingga seluruh cabang mengarah pada kelas target yang jelas (stres ringan, sedang, atau berat).

Model C4.5 di bangun menggunakan 70 data training dan di testing menggunakan 30 data testing. Selanjutnya dilakukan pengujian dengan menguji dataset dengan algoritma C4.5 untuk mengukur akurasi prediksi dari klasifikasi dataset. Pengujian dataset dengan algoritma C4.5 menggunakan Confusion Matrix dengan hasil seperti Tabel 6 di bawah ini.

Tabel 6. Confusion Matrix Algoritma C4.5

		Predicted Values		
		0	1	2
Actual Values	0	104	8	1
	1	5	98	4
	2	9	4	97

3.2.2 Algoritma SVM

Model SVM dengan kernel RBF menunjukkan konvergensi yang baik setelah 70 iterasi training dan 30 iterasi testing. Parameter C=1.0 memberikan keseimbangan optimal antara margin maximization dan error tolerance. Hasil dari model klasifikasi SVM dapat dilihat pada Tabel 7.

Tabel 7. Hasil Prediksi SVM

Blood Pressure	Self Esteem	Sleep Quality	Lecturer Student Relationship	Bullying	Actual Test	Prediction Test
3	2,5	1	5	5	2	2
2	4,3	5	1	1	0	0
1	2,5	2	2	2	1	1
1	4	2	2	3	1	1
2	4,3	4	1	1	0	0

Berdasarkan tabel di atas dapat dilihat bahwa Actual test dan Prediction_test terdapat kesalahan dalam memprediksi Stress Level. Kesalahan dalam Memprediksi dapat dilihat pada Confusion Matrix. Tabel Confusion Matrix dapat dilihat pada Tabel 8.

Tabel 8. Confusion Matrix SVM

	Predicted Values			
	0	1	2	
Actual Values	0	103	5	5
	1	4	98	5
	2	7	5	98

3.3 Evaluasi Model SVM dan C4.5

Evaluasi model dapat dilihat pada matrix evaluasi. Matriks ini membantu dalam memahami sejauh mana model SVM mampu memberikan hasil klasifikasi yang akurat dan andal. Berdasarkan perhitungan maka didapatkan hasil matriks evaluasi seperti yang di lihatkan pada Tabel 9 di bawah ini.

Tabel 9. Matriks Evaluasi C4.5

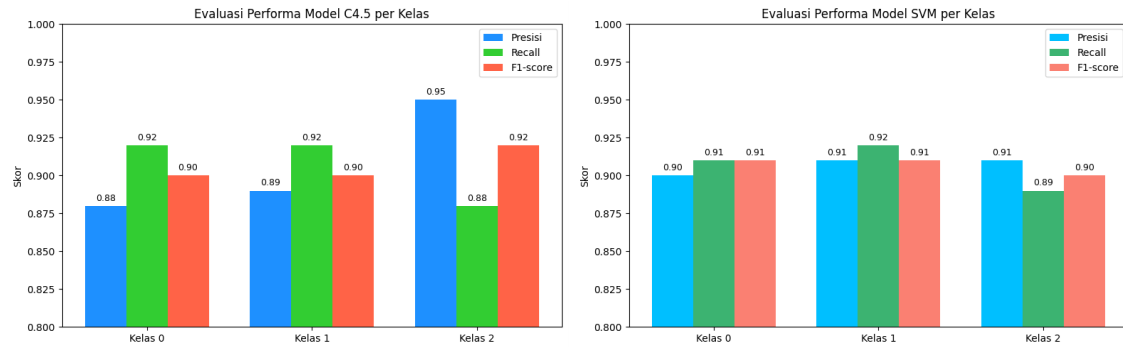
Matriks Evaluasi	0	1	2
Presisi	88%	89%	95%
Recall	92%	92%	88%
F1-score	90%	90%	91%
Akurasi	91%		

Hasil evaluasi menunjukkan bahwa algoritma C4.5 mencapai akurasi sebesar 91% dengan nilai presisi tertinggi pada kelas stres berat (95%), sedangkan nilai recall tertinggi terdapat pada kelas stres ringan dan sedang (92%). Hal ini menunjukkan bahwa algoritma C4.5 cukup efektif dalam mengenali semua kelas tingkat stres mahasiswa dengan performa yang stabil di semua metrik evaluasi. Namun, perbandingan dengan algoritma SVM yang memiliki akurasi serupa (91%) mengindikasikan bahwa kedua metode dapat digunakan secara efektif, meskipun SVM cenderung lebih unggul dalam kestabilan prediksi.

Tabel 10. Evaluasi Matriks SVM

Matriks Evaluasi	0	1	2
Presisi	90%	91%	91%
Recall	91%	92%	91%
F1-score	91%	91%	90%
Akurasi	91%		

Model klasifikasi memiliki performa yang cukup baik dan konsisten untuk ketiga kelas (0, 1, dan 2). Presisi tinggi (90–91%) menunjukkan bahwa mayoritas prediksi model untuk tiap kelas adalah benar, artinya jumlah kesalahan prediksi (*false positive*) relatif rendah. Recall yang juga tinggi (91–92%) menunjukkan bahwa model berhasil menangkap sebagian besar data yang benar-benar termasuk dalam masing-masing kelas, dengan kelas 1 memiliki recall tertinggi (92%), yang berarti data kelas 1 paling sedikit terlewat (*false negative* rendah). F1-score yang mendekati nilai 1 (atau 100%) menandakan keseimbangan yang baik antara presisi dan recall di ketiga kelas. Meskipun F1 kelas 2 sedikit lebih rendah (90%), perbedaannya tidak signifikan dan masih tergolong tinggi. Akurasi keseluruhan model adalah 91%, yang berarti 91% dari seluruh data berhasil diklasifikasikan dengan benar. Ini mencerminkan bahwa model cukup andal dan memiliki kinerja yang merata di semua kelas, tanpa dominasi atau bias terhadap satu kelas tertentu.



Gambar 3. Komparasi Model C4.5 dan SVM

Kestabilan kinerja SVM didukung oleh tahapan praproses yang mencakup MinMax Scaling dan seleksi fitur menggunakan Backward Elimination, yang membantu mengurangi atribut tidak relevan dan memastikan distribusi fitur seragam. Di sisi lain, performa C4.5 yang sedikit lebih rendah kemungkinan disebabkan oleh sifat pohon keputusan yang rentan terhadap overfitting pada dataset yang kompleks, meskipun perhitungan manual entropy dan gain telah menunjukkan pemilihan atribut yang optimal.

Secara keseluruhan, kedua algoritma layak digunakan dalam klasifikasi tingkat stres mahasiswa. Namun, SVM lebih direkomendasikan untuk implementasi nyata karena konsistensi hasil, ketahanan terhadap variasi data, serta akurasi yang relatif lebih stabil. Selain itu, penelitian ini menunjukkan bahwa penerapan teknik seleksi fitur (Backward Elimination) memberikan kontribusi signifikan terhadap peningkatan performa kedua algoritma, yang terlihat dari penyempitan atribut menjadi lima variabel utama dengan signifikansi yang tinggi.

4. Simpulan

Penelitian ini berfokus pada komparasi algoritma *Support Vector Machine* (SVM) dan C4.5 dalam klasifikasi tingkat stres mahasiswa dengan memanfaatkan teknik praproses *MinMax Scaling* dan seleksi fitur *Backward Elimination*. Dengan menggunakan dataset dari Kaggle yang terdiri atas 1.100 data mahasiswa, penelitian ini berhasil mengeksplorasi efektivitas kedua metode tersebut dalam mendeteksi tingkat stres yang dikategorikan ke dalam tiga kelas: stres ringan, sedang, dan berat. Hasil penelitian menunjukkan bahwa kedua algoritma memiliki akurasi yang sama, yakni 91%, tetapi terdapat perbedaan dalam distribusi nilai metrik evaluasi lainnya. SVM menunjukkan presisi, recall, dan F1-score yang lebih stabil di seluruh kategori dibandingkan C4.5, menegaskan keunggulan SVM dalam menangani data berdimensi tinggi yang telah melalui proses normalisasi.

Sebagai rekomendasi, penelitian selanjutnya dapat memperluas cakupan dengan menambahkan algoritma berbasis ensemble learning seperti Random Forest atau Gradient Boosting, serta mengintegrasikan data fisiologis atau *sensor real-time* untuk memperkaya fitur input dan meningkatkan akurasi prediksi. Selain itu, penerapan model ini pada platform berbasis web atau aplikasi seluler akan memungkinkan penggunaannya secara langsung oleh mahasiswa atau pihak kampus sebagai alat deteksi dini. Dengan demikian, penelitian ini tidak hanya memberikan kontribusi teoritis dalam bidang *data mining* dan *machine learning*, tetapi juga nilai praktis yang dapat mendukung upaya peningkatan kesehatan mental mahasiswa secara lebih sistematis.

Pustaka

- [1] S. A. Farika, M. N. Mirza, and A. N. Romas, "Promosi kesehatan tentang pentingnya menjaga kesehatan mental pada remaja," *Jurnal Pengabdian dan Pemberdayaan Kesehatan*, vol. 1, no. 1, pp. 69–77, Jun 2024.
- [2] H. D. Putra, L. Khairani, D. Hastari, P. S. S. Informasi, F. S. dan Teknologi, and C. Author, "Comparison of naive bayes classifier and support vector machine algorithms for classifying student mental health data," in *SENTIMAS Seminar Nasional Penelitian dan Pengabdian Masyarakat*, 2023, pp. 120–125.

- [3] N. F. Mahing, A. L. Gunawan, A. F. A. Zen, F. A. Bachtiar, and S. A. Wicaksono, "Klasifikasi tingkat stress dari data berbentuk teks dengan menggunakan algoritma support vector machine (svm) dan random forest," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 5, pp. 1067–1076, Oct 2024.
- [4] W. Kurniawati and R. Setyaningsih, "Manajemen stress pada mahasiswa tingkat akhir dalam penyusunan skripsi," *Jurnal An-Nur Kajian Pendidikan dan Ilmu Keislaman*, vol. 5, no. 3, pp. 248–253, 2020.
- [5] M. Sutjiato and G. D. K. a a T. Tucunan, "Hubungan faktor internal dan eksternal dengan tingkat stress pada mahasiswa fakultas kedokteran universitas sam ratulangi manado," *JIKMU*, vol. 5, no. 1, pp. 30–42, 2015.
- [6] S. Bahri, D. M. Midyanti, R. Hidayati, J. S. Komputer, and F. Mipa, "Perbandingan algoritma naive bayes dan c4.5 untuk klasifikasi penyakit anak," in *Seminar Nasional Aplikasi Teknologi Informasi*, 2018, pp. 11–2018.
- [7] N. W. S. Agustini, D. Priadi, and R. V. Atika, "Profil kimia dan aktivitas antibakteri fraksi aktif nan-nochloropsis sp. sebagai senyawa penghambat bakteri penyebab gangguan kesehatan mulut," *Jurnal Pascapanen dan Bioteknologi Kelautan dan Perikanan*, vol. 17, no. 1, p. 19, 2022.
- [8] S. F. Damanik, A. Wanto, and I. Gunawan, "Penerapan algoritma decision tree c4.5 untuk klasifikasi tingkat kesejahteraan keluarga pada desa tiga dolok," *Jurnal Krisnadana*, vol. 1, no. 2, pp. 21–32, 2022.
- [9] G. Abdurrahman, "Klasifikasi kanker payudara menggunakan algoritma svm dengan kernel rbf, linier, dan sigmoid," *JUSTIFY: Jurnal Sistem Informasi Ibrahimy*, vol. 2, no. 1, pp. 74–80, 2023.
- [10] I. A. Ropikoh, R. Abdulhakim, U. Enri, and N. Sulistiyowati, "Penerapan algoritma support vector machine (svm) untuk klasifikasi berita hoax covid-19," *Journal of Applied Informatics and Computing*, vol. 5, no. 1, pp. 64–73, Jul 2021.
- [11] U. M. Pratiwi and M. Ibad, "Klasifikasi faktor yang berpengaruh dalam kehamilan tidak diinginkan menggunakan metode algoritma decision tree," *Jurnal Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika dan Statistika*, vol. 3, no. 2, pp. 406–416, 2022.
- [12] Y. Anggraiwan and B. Siregar, "Klasifikasi harga mobil menggunakan metode decision tree algoritma c4.5," *COMPUTATIO: Journal of Computer Science and Information Systems*, vol. 6, no. 2, pp. 70–79, 2022.
- [13] A. A. Yaqin, M. A. Barata, and N. Mahmudah, "Implementation of the random forest algorithm with optuna optimization in lung cancer classification," *SISTEMASI*, vol. 14, no. 2, pp. 561–569, 2025.
- [14] P. P. Allorerung, A. Erna, M. Bagussahrir, and S. Alam, "Analisis performa normalisasi data untuk klasifikasi k-nearest neighbor pada dataset penyakit," *JISKA (Jurnal Informatika Sunan Kalijaga)*, vol. 9, no. 3, pp. 178–191, 2024.
- [15] M. Ilmiyah, M. A. Barata, and P. E. Yuwita, "Implementation of ann optimization with smote and backward elimination for pcps prediction," *Scientific Journal of Informatics*, vol. 12, no. 1, pp. 133–144, 2025.
- [16] A. L. Rahmiyati, S. Anwar, G. Irianto, and A. D. Abdillah, "Sistem manajemen logistik obat di instalasi farmasi puskesmas ciparay," *Jurnal Kesehatan Kartika*, vol. 17, no. 1, pp. 28–33, 2022.
- [17] E. F. Laili, Z. Alawi, R. Rohmah, and M. A. Barata, "Komparasi algoritma decision tree dan support vector machine (svm) dalam klasifikasi serangan jantung," *Jurnal Sistem Informasi dan Informatika*, vol. 8, no. 1, pp. 67–76, 2025.
- [18] M. A. Barata, E. Noersasongko, Purwanto, and M. A. Soeleman, "Improving the accuracy of c4.5 algorithm with chi-square method on pure tea classification using electronic nose," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 2, pp. 226–235, 2023.
- [19] C. Anam and H. B. Santoso, "Perbandingan kinerja algoritma c4.5 dan naive bayes untuk klasifikasi penerima beasiswa," *Jurnal Energy*, vol. 8, no. 1, pp. 2088–4591, 2018.

- [20] R. Resmiati and T. Arifin, "Klasifikasi pasien kanker payudara menggunakan metode support vector machine dengan backward elimination," *Sistemasi*, vol. 10, no. 2, p. 381, 2021.
- [21] F. Ma'arif and T. Arifin, "Optimasi fitur menggunakan backward elimination dan algoritma svm untuk klasifikasi kanker payudara," *Jurnal Informatika*, vol. 4, no. 1, pp. 46–53, 2017.
- [22] A. Ambarwari, Q. J. Adrian, and Y. Herdiyeni, "Analisis pengaruh data scaling terhadap performa algoritme machine learning untuk identifikasi tanaman," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 4, no. 1, pp. 117–122, 2020.