**Journal of Applied Informatics and Computing (JAIC)**
Vol.9, No.3, June 2025, pp. 633~641
e-ISSN: 2548-6861


# Optimization of *Random Forest* Algorithm with *Backward Elimination* Method in Classification of Academic Stress Levels

**Salsabila Dani Amalia [1]*, Mula Agung Barata [2]*, Pelangi Eka Yuwita [3]****
* Teknik Informatika, Universitas Nahdlatul Ulama Sunan Giri
**Teknik Mesin, Universitas Nahdlatul Ulama Sunan Giri
salsabilladani28@gmail.com [1]

**Article Info**

***Article history:***

Received 2025-03-14
Revised 2025-03-21
Accepted 2025-03-26

***Keyword:***

*Academic stress,*
*Classification,*
*Random Forest,*
*Backward Elimination.*

**ABSTRACT**

Stress is a phenomenon experienced by all individuals as a natural response to pressure, which can impact mental and physical health. In an academic setting, the stress experienced by students is known as academic stress, which can affect their performance and mental well-being. Therefore, there is a need for effective prediction methods to aid in the management and prevention of academic stress. Therefore, there is a need to predict the level of academic stress to aid more effective management and prevention. This study uses a public dataset categorized based on the Student-life Stress Inventory (SSI), which includes psychological, physiological, social, environmental, and academic factors. Data mining is often used to detect diseases, one of which is the Random Forest algorithm. The Random Forest algorithm is applied as a classification technique for academic stress levels, with optimization using the Backward Elimination method for feature selection to improve model accuracy. The results showed that the accuracy of the Random Forest algorithm without feature selection obtained an accuracy of 86%, compared to the random forest algorithm with feature selection using the Backward Elimination method obtained a higher accuracy of 88%. This increase shows that the feature selection method can optimize model performance by selecting more relevant features. Thus, this research is expected to contribute to the management of student academic stress against the risk of academic stress.



## I. PENDAHULUAN

Kondisi mental yang paling umum yang sulit dicegah salah satunya yakni stres, yang berdampak buruk pada kemampuan seseorang untuk berpikir jernih dan bertindak secara normal. Stres dapat di artikan sebagai respons fisik terhadap tantangan mental, emosional, dan fisik[1]. Ketika kita dihadapkan pada suatu masalah atau berada dalam situasi yang menuntut kita untuk beradaptasi dengan cepat terhadap suatu perubahan, kita sedang mengalami stres fisik[2]. Menurut laporan WHO tahun 2018, 350 juta orang di seluruh dunia menderita stres, di tingkat global, prevelasi stress siswa mencapai 38,91%, dengan angka yang lebih tinggi di Asia (61,3%) dan di Indonesia (71,6%). Sementara itu, Kementerian Kesehatan melaporkan bahwa lebih dari 19 juta penduduk Indonesia yang berusia di atas 15 tahun menderita penyakit mental dan emosional[3]. Meskipun stres merupakan hal yang normal dan sering ditemui dalam kehidupan, stres juga dapat muncul dalam lingkungan akademik [4].

Salah satu masalah kesehatan terbesar yang memengaruhi lingkungan akademik disebut dengan stres akademik[5]. Stres akademik menunjukkan masalah kesehatan yang berdampak sangat besar pada prestasi akademik. Stres akademik yang biasanya dialami oleh siswa atau mahasiswa disebabkan oleh kurangnya semangat akademik[3]. Menurut tingkatannya, stress dibagi menjadi tiga kategori yaitu stres ringan, stres sedang, dan stres berat[6]. Stres yang dialami dilingkungan akademik beragam mulai dari tingkat sedang hingga berat yang dialami ini dapat menghambat proses belajar mengajar[7]. Ciri-ciri stres akademik adalah lelah, cemas, kesulitan berkonsentrasi dan kurangnya kontrol diri. Akibatnya, tugas menjadi tertunda, merasakan putus asa dan

kehilangan minat belajar[8]. Oleh karena itu, penting bagi seorang siswa dan mahasiswa untuk memiliki pengelolahan stress yang baik. Pengelolaan stres yang baik membuat mahasiswa terhindar dari dampak buruk yang ditimbulkan oleh stres.

Dalam penelitian ini, data mengenai stres dianalisis menggunakan pendekatan *Studen-life Stress Inventory* (SSI). SSI merupakan media yang digunakan untuk mengukur berbagai faktor yang berkonstribusi terhadap stres siswa, termasuk faktor psikolog, fisiologis, social, lingkungan, dan akademik[9]. Dataset yang digunakan dalam penelitian ini telah dikategorikan berdasarkan lima faktor utama yang relevan dengan model SSI, yang mencakup variabel seperti tingkat kecemasan, kualitas tidur, tekanan akademik, serta interaksi social. Dengan demikian, dataset ini memiliki kesamaan dengan metode pengukuran dengan SSI yang telah banyak digunakan dalam penelitian terkai stress akademik.

Dalam situasi ini, pengklasifikasian mengenai stres sangat penting untuk membantu dalam pengelolahan stres, pencegahan masalah kesehatan dan meningkatkan kualitas hidup. Untuk mengatasi hal ini, institusi pendidikan perlu memprediksi resiko stres akademik dengan menganalisis data mahasiswa dan menemukan pola perilaku yang relevan. Salah satu pendekatan yang sering digunakan adalah data mining, klasifikasi termasuk salah satu algoritma data mining yang memiliki cara pengelompokan benda berdasarkan ciri - ciri yang dimiliki oleh objek klasifikasi[10]. Dalam penelitian ini, Teknik klasifikasi diterapkan untuk memproses data latih dan data uji, serta menilai akurasi model dalam memprediksi stres akademik.

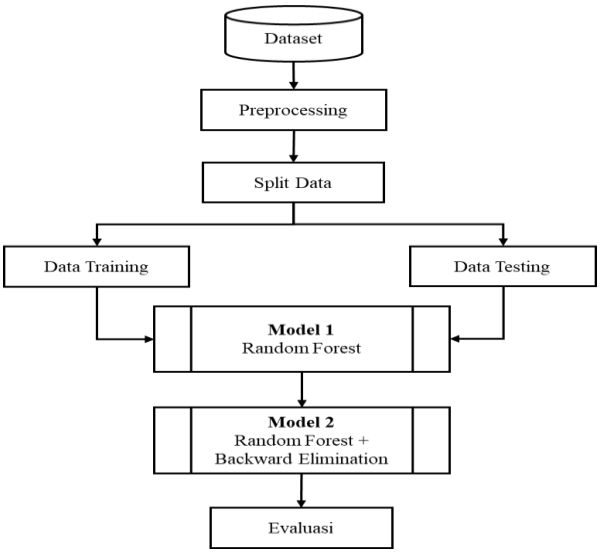
Penelitian terkait stres telah banyak dilakukan dengan berbagai pedekatan metode. Sebayang, dkk menggunakan algoritma *Random Forest* dalam menguji dataset kesehatan mental memproleh hasil akurasi sebesar 84%[11]. Kemudian penelitian yang dilakukan Mohamed, dkk pada penelitiannya, menggunakan dataset kesehatan mental. Dengan komparasi tiga metode yaitu *Support Vector Machine* (SVM), *Random Forest* dan *Multilayer Perception* (MPL) diperoleh hasil akurasi paling tinggi sebesar 97.67% dari algoritma *Random Forest*[12]. Ini menunjukkan betapa efektifnya *Random Forest* dalam melakukan pengelompokan dengan hasil yang akurat. Namun metode ini memiliki kelemahan yakni sulit memilih fitur yang relevan dan optimal pada nilai atribut yang digunakan akibatnya nilai akurasi menjadi kecil.

Salah satu metode yang dapat meningkatkan kinerja algoritma adalah algoritma *Backward Elimination,* digunakan untuk melakukan seleksi fitur guna mengoptimalkan model klasifikasi. Menurut M. Rudi, dkk dalam penelitiannya mengenai analisis strategi penjuaalan produk UMKM dengan hasil akurasi algoritma C4.5 sebesar 82,78% namun setelah menerapkan *Backword Elimination* mengalami peningkatan akurasi sebesar 85.00%[13]. Adapun penelitian yang dilakukan Cahyaningtyas, dkk melakukan Perbandingan algoritma dan fitur seleksi untuk klasifikasi hasil panen ayam broiler, menggunakan algoritma *Naïve bayes*, *Random Forest*, K-NN, *Forward Selection* dan *Backward Elimination* memperoleh hasil dari perbandingan algoritma *Random Forest* menjadi algoritma terbaik dengan akurasi sebesar 89,14%, sedangkan RF dikombinasikan dengan BE menghasilkan akurasi tertinggi sebesar 96,67%[14]. Oleh sebab itu, penerapan teknik seleksi fitur *Backward Elimination* dapat menjadi solusi untuk meningkatkan akurasi model dalam mengklasifikasikan tingkat stress akademik.

Berdasarkan uraian penelitian diatas, dapat disimpulkan bahwa pemodelan prediksi stress akademik dengan metode klasifikasi memerlukan pendekatan yang tepat dalam pemilihan algoritma dan seleksi fitur. Oleh karena itu, dalam penelitian ini memutuskan mengunakan algoritma *Random Forest* dengan metode *Backward Elimination* dalam klasifikasi tingkat stress akademik. Dengan demikian, diharapkan hasil penelitian dapat memberikan konstribusi dalam membantu mahasiswa untuk mengidentifikasi faktor resiko stress yang mereka hadapi.

## II. METODE

Proses skema penelitian dapat dilihat pada gambar 1, yang menggambarkan berbagai bagian dan tahapan prosedur, memberikan gambaran umum tentang rencana yang disarankan. Penggambaran grafis ini memudahkan untuk memahami bagaimana berbagai komponen bekerja bersama dan mendukung tujuan utama penelitian.




Gambar 1 alur metode

### A. Dataset

Pada penelitian ini dataset yang digunakan merupakan data publik yang didapatkan dari website *Kaggle.com*[15], dataset terdiri dari 20 atribut dengan perilaku yang mempengaruhi tingkat stres dan 1 label terdapat 3 kelas yakni kelas stres ringan, sedang dan berat, yang merupakan hasil penjumlahan angka perilaku yang condong pada tingkatan stres akademik dengan jumlah dataset 1100 data. Atribut dataset dapat di lihat pada tabel 1 sebagai berikut.

TABEL 1
ATRIBUT DATASET

| No | Atribut | Nilai | Deskripsi |
|---|---|---|---|
| 1. | *Anxiety level* | 0 – 21 | Nilai yang lebih tinggi menunjukkan tingkat kecemasan tinggi, diukur dengan GAD-7 |
| 2. | *Self esteem* | 0 – 30 | Nilai yang lebih tinggi menunjukkan hargadiri yang lebih baik, di ukur dengan *Rosenberg Self Esteem Scale* |
| 3. | *Mental health history* | 0 dan 1 | 0 = tidak ada riwayat dan 1 = ada riwayat |
| 4. | *Depression* | 0 – 27 | Nilai yang tinggi menunjukkan tingkat depresi yang lebih tinggi, diukur dengan PHQ-9 |
| 5. | *Headache* | 1 – 5 | Nilai Sakit kepala yang dirasakan 0-1 = Rendah, 2-3, Menengah, dan 4-5 Tinggi |
| 6. | *Blood pressure* | 1 – 5 | Nilai tekanan darah 0-1 = Rendah, 2-3, Menengah, dan 4-5 Tinggi |
| 7. | *Sleep quality* | 1 – 5 | Nilai kualitas tidur 0-1 = Rendah, 2-3, Menengah, dan 4-5 Tinggi |
| 8. | *Breathing problem* | 1 – 5 | Nilai masalah pernapasan 0-1 = Rendah, 2-3, Menengah, dan 4-5 Tinggi |
| 9. | *Noise level* | 1 – 5 | Nilai tingkat kebisingan 0-1 = Rendah, 2-3, Menengah, dan 4-5 Tinggi |
| 10. | *Living conditions* | 1 – 5 | Nilai kondisi tempat tinggal 0-1 = Rendah, 2-3, Menengah, dan 4-5 Tinggi |
| 11. | *Safety* | 1 – 5 | Nilai rasa keamanan 0-1 = Rendah, 2-3, Menengah, dan 4-5 Tinggi |
| 12. | *Basic needs* | 1 – 5 | Nilai kebutuhan dasar 0-1 = Rendah, 2-3, Menengah, dan 4-5 Tinggi |
| 13. | *Academic performance* | 1 – 5 | Nilai performa akademis 0-1 = Rendah, 2-3, Menengah, dan 4-5 Tinggi |
| 14. | *Study load* | 1 – 5 | Nilai beban belajar 0-1 = Rendah, 2-3, Menengah, dan 4-5 Tinggi |
| 15. | *Teacher student relationship* | 1 – 5 | Nilai hubungan guru-murid 0-1 = Rendah, 2-3, Menengah, dan 4-5 Tinggi |
| 16. | *Peer pressure* | 1 – 5 | Nilai tekanan teman sebaya 0-1 = Rendah, 2-3, Menengah, dan 4-5 Tinggi |
| 17. | *Future career concerns* | 1 – 5 | Nilai kekhawatiran akan karir di masa depan 0-1 = Rendah, 2-3, Menengah, dan 4-5 Tinggi |
| 18. | *Social support* | 1 – 5 | Nilai dukungan sosial 0-1 = Rendah, 2-3, Menengah, dan 4-5 Tinggi |
| 19. | *Extracurricular activities* | 1 – 5 | Nilai kegiatan ekstrakurikuler 0-1 = Rendah, 2-3, Menengah, dan 4-5 Tinggi |
| 20. | *Bullying.* | 1 – 5 | Nilai perundungan 0-1 = Rendah, 2-3, Menengah, dan 4-5 Tinggi |

### *B. Preprocessing*

*Preprocessing* merupakan serangkaian langkah untuk mempersiapkan data sebelum dilakukan proses klasifikasi, guna menjamin kualitas dan ketepatan hasil analisis, prosedur ini sangat penting dalam bidang data mining dan *machine learning*[16]. Prosedur ini terdiri dari beberapa langkah penting, termasuk reduksi data untuk membuat data menjadi lebih sederhana tanpa menghilangkan informasi penting, transformasi data untuk mengubah format data sesuai dengan kebutuhan analisis, integrasi data untuk menggabungkan informasi dari berbagai sumber, dan pembersihan data untuk mengatasi data yang hilang atau tidak konsisten. Menerapkan persiapan data sebaik mungkin dapat meningkatkan kualitas data yang digunakan untuk analisis, menghasilkan informasi yang lebih akurat dan dapat dipercaya.

### *C. Algoritma Random Forest*

*Random Forest* merupakan salah satu algoritma klasifikasi dan regresi dalam data mining dan *machine learning*. *Random Forest* menggabungkan konsep dari metode *Decision Tree* dan *ensemble*[11]. *Random Forest* didasarkan pada algoritma bagging algoritma dan menggunakan teknik *Ensemble Learning* di mana membuat pohon sebanyak mungkin pada subset data dan menggabungkan output dari semua pohon. Sebagai hasilnya, algoritma *Random Forest* mencapai pengurangan masalah *overfitting* pada pohon keputusan dan juga pengurangan varian, yang pada akhirnya meningkatkan akurasi[16].

*Random Forest* digunakan dalam klasifikasi dan regresi linier dan bekerja dengan baik dengan variabel kategorikal dan variabel. Ini menggunakan pendekatan berbasis aturan dari perhitungan jarak dan sebagai hasilnya tidak diperlukan penskalaan fitur standarisasi dan parameter nonlinier tidak mempengaruhi kinerja *Random Forest* tidak seperti berbasis kurva algoritma dan sangat stabil dan relatif lebih sedikit terpengaruh oleh noise[17].

Pada algoritma *Random Forest*, *IndexGini* dan *Gini Gain*, yang bertindak sebagai kriteria pengukur ketidak pastian dalam data saat melakukan pemisahan, yang digunakan dalam algoritma *Random Forest* untuk mengevaluasi kualitas pemisahan data. Nilai *IndexGini* dapat dihitung menggunaan persamaan (2) sebagai berikut:

$$IndexGini(S) = 1 - \sum_{i=1}^{k} (Pi^2) \qquad (2)$$

Keterangan:

*IndexGini(S)* = Sekumpulan data
Pi = Proporsi data pada kelas *i*
*K* = Jumlah kelas

*IndexGini* membantu memilih fitur optimal untuk setiap pemisahan dengan meminimalkan ketidakmurnian, yang menghasilkan pohon yang lebih efisien yang memisahkan data dengan lebih sukses. Semakin rendah nilai gini, maka semakin baik artinya data lebih homogen. Selanjutnya, *GiniGain* digunakan untuk menilai kualitas pemisahan menggunakan rata-rata tertimbang dari Gini sub-set yang

dihasilkan. Formula dapat digunakan untuk menentukan nilai *GiniGain* dengan menggunakan persamaan (3):

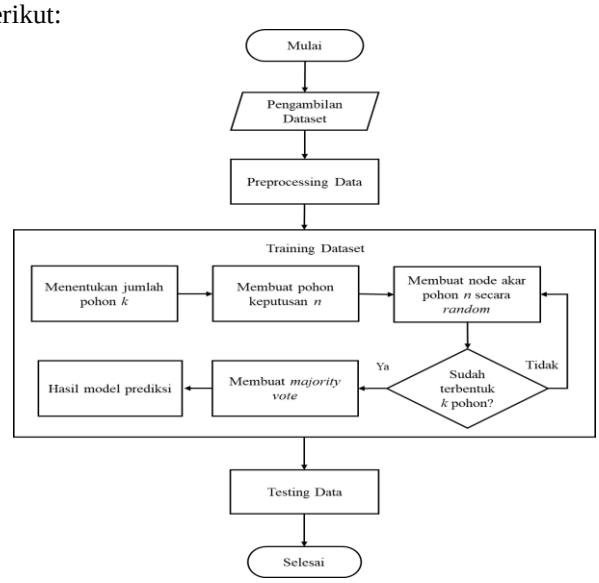
$$GiniGain(A,S) = Gini(S) - \sum_{k=1}^{k} \frac{|Si|}{|S|} * Gini(S_i) \quad (3)$$

Keterangan:

Gini(S) = indeks gini dari dataset sebelum dipisah
A = atribut yang digunakan untuk membagi dataset
Si = Jumlah sampel dalam subset Si setelah pemisahan berdasarkan atribut A
Gini(Si) = Nilai Gini dari subset (Si)

*GiniGain* sangat pentung untuk mengukur seberapa besar penurunan ketidakmurnian yang signifikan setelah pemisahan berdasarkan atribut A semakin tinggi *GiniGain*, semakin baik atribut tersebut dalam menganalisis data. Kedua formula ini sangat penting dalam proses pembelajaran *Random Forest* karena *IndexGini* membantu menentukan seberapa baik sebuah fitur dapat memisahkan data. Fitur yang optimal berdasarkan penurunan impuritas dipilih dengan menggunakan *GiniGain* dan fitur optimal berdasarkan penurunan impuritas dipilih dengan menggunakan *IndexGini* berdasarkan penurunan ketidakmurnian yang dihasilkan. Dengan menggabungkan output dari banyak pohon keputusan, *Random Forest* dapat menghasilkan prediksi yang lebih akurat.

Penerapan *Random Forest* dalam melakukan proses klasifikasi menunjukkan kinerja yang memuaskan menurut penelitian terdahulu yang sudah pernah dilakukan. Hasil kinerja metode *Random Forest* yang baik tersebut melalui langkah-langkah yang sistematis sesuai dengan referensi yang disajikan oleh penelitian terdahulu adapun tahap dalam perhitungan *Random Forest* pada gambar 2 adalah sebagai berikut:




Gambar 2 Alur Perhitungan Random Forest

*D. Backward Elimination*

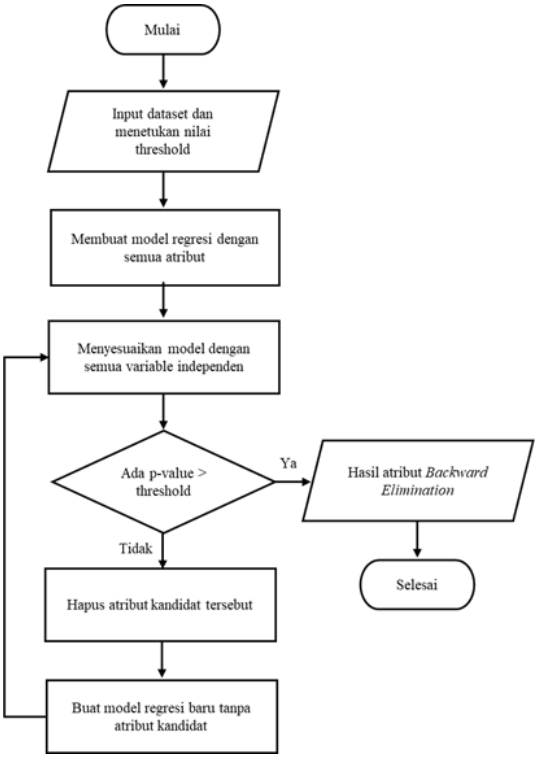
Metode *Backward Elimination* mampu menghapus fitur yang tidak diperlukan dan hanya menggunakan fitur yang memenuhi batas threshold pada parameter yang telah ditetapkan sebelumnya[15]. *Backward Elimination* bekerja dengan cara mengevaluasi seluruh fitur, kemudian secara bertahap mengeliminasi fitur yang tidak signifikan, proses ini di ulang hingga semua variabel yang tersisa dalam model memiliki nilai yang signifikan. Tingkat signifikan menentukan pemilihan atribut, dimana semakin kecil nilai signifikan level, maka proses seleksi atribut lebih ketat, sehingga hanya sedikit atribut yang akan dipilih[16][17]. Adapun rumus dalam perhitungan Backward Elimination menggunakan persamaan (1) adalah sebagai berikut:

$$y = b_0 + b_1 * x_1 + b_2 * x_2 + \cdots + b_n * x_n \quad (1)$$

Keteranagan:

$Y$ = Variabel terikat
$X$ = Variabel bebas
$b_n$ = koefisien regresi

Penerapan *Backward Elimination* dalam melakukan proses *feature selection* menunjukkan kinerja yang memuaskan menurut penelitian terdahulu yang sudah pernah dilakukan. Hasil kinerja metode *Backward Elimination* yang baik tersebut melalui langkah-langkah yang sistematis sesuai dengan referensi yang disajikan oleh penelitian terdahulu adapun tahap dalam perhitungan *Backward Elimination* pada gambar 3 adalah sebagai berikut:




Gambar 3 Alur Perhitungan Backward Elimination

*E. Evaluasi dan Validation*

Tahap selanjutnya adalah melakukan analisis dan mengevaluasi hasil klasifikasi kinerja *Random Fores* dan mengevaluasi hasil yang didapatakan sebagai bahan untuk mempelajari bagaimana algoritma *Random Forest* ini dapat meningkatkan akurasinya pada dataset tingkat stres akademik. *Confusion Matrix* merupakan salah satu alat yang digunakan untuk evaluasi pengujian yakni tabel yang menggambarkan hasil prediksi model terhadap data sebenarnya. Ini terdiri dari empat sel: *True positive* (TP), *True Negative* (TN), *False Positive* (FP) dan *False Negative* (FN)[18]. Proses pembentukannya melalui training model dan prediksi, kemudian semua sampel di uji dan dibuat tabulasi jumlah sampel yang di klasifikasi. Berdasarkan hasil klasifiksi model pada data uji jika akurasi model tinggi akan berada di diagonal utama, jika banyak kesalahan akan ada nilai besar di sel lain yakni FP atau FN.

Nilai efektivitas dari keseluruhan kinerja model merupakan nilai akurasi[19]. Nilai akurasi menampilkan proporsi klasifikasi yang akurat di antara semua pengamatan data penelitian. Persamaan (4) menunjukkan bagaimana nilai akurasi di peroleh.

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \qquad (4)$$

Nilai presisi digunakan untuk menunjukkan kategorisasi nilai positif dibandingkan dengan perkiraan nilai positif dan untuk mengukur keakuratan kinerja model klasifikasi[19]. persamaan (5) menunjukkan bagaimana presisi dihitung.

$$Precision = \frac{TP}{TP + FP} \qquad (5)$$

Nilai recall merupakan nilai rata-rata *true positive* untuk menghitung seberapa baik model dalam mengidentifikasi kelas normal[19]. Dalam mengidentifikasi kelas normal, pada persamaan (6) merupakan cara memperoleh nilai *recall*.

$$Recall = \frac{TP}{FN + TP} \qquad (6)$$

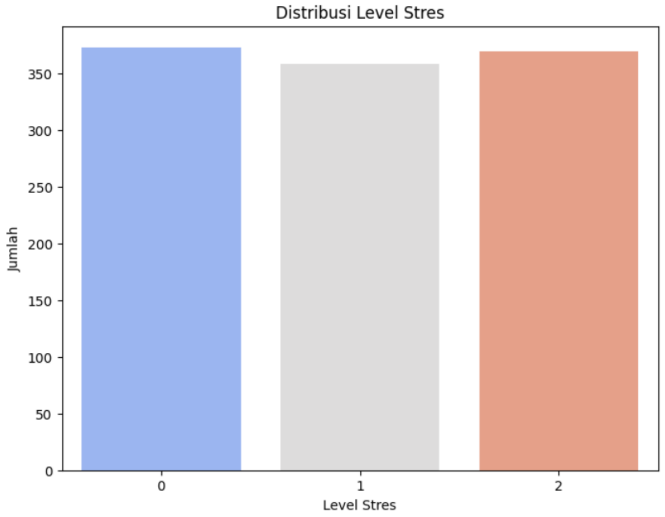
## III. HASIL DAN PEMBAHASAN

Pada tahap ini berisi tentang penjelasan mengenai hasil analisin dan kombinasi algoritma *Random Forest* dan *Backward Elimination*, sebagai analisis pemilihan fitur yang relevan, dalam memprediksi tingkat stres akademik.

Dataset tingkat stres akademik merupakan data publik dipeoleh dari website *Kaggle.com*. Dataset ini berjumlah 1100 data, terdapat 21 atribut, memiliki 3 kelas yaitu 0 sebagai stress ringan, 1 stres sedang dan 2 stres berat.

TABEL 2
DATASET TINGKAT STRES AKADEMIK

| No | Anxiety level | Self esteem | Mental health history | Depression | … | Stress Level |
|---|---|---|---|---|---|---|
| **1** | 14 | 20 | 0 | 11 | … | 1 |
| **2** | 15 | 8 | 1 | 15 | … | 2 |
| **3** | 12 | 18 | 1 | 14 | … | 1 |
| **4** | 16 | 12 | 1 | 15 | … | 2 |
| **5** | 16 | 28 | 0 | 7 | … | 1 |
| **…** | … | … | … | … | … | … |
| **…** | … | … | … | … | … | … |
| **…** | … | … | … | … | … | … |
| **1100** | 18 | 6 | 1 | 15 | … | 2 |




Gambar 4 Visualisasi data Level Stres

Pada gambar 4 menunjukkan distribusi level stres dalam 3 kategori (0, 1, dan 2). Dari grafik tersebut, dapat dilihat bahwa jumlah dalam setiap kategori relative sama, tanpa perbedaan yang signifikan. Hal ini mengindikasikan bahwa data target bersifat seimbang, sehinggan tidak diperlukan teknik penyeimbangan data seperti *oversampling* atau *undersampling*. Dengan demikian, proses *preprocessing* seperti normalisasi ulang atau penanganan ketidak seimbangan data tidak diperlukan, dan model dapat dilihat secara langsung tanpa khawatir terhadap bias distribusi kelas.

*A. Preprocessing Data*

Preprocessing merupakan tahap awal dalam analisis data yang bertujuan untuk meningkatkan kualitas data dan sudah siap uji dengan algoritma yang dipilih dengan melakukan pengamatan apakah data yang akan di olah mengandung *missing value, noise, atau outlier*.

*1) Handling Missing Value*

Pada tahap ini, digunakan untuk mengecek apakah pada dataset terdapat *missing value*, jika terdapat nilai kosong maka akan dilakukan cleaning data, namun pada gambar 5 menunjukkan bahwa pada dataset tidak terdapat missing value.

```
print(data.isnull().sum())
```

```
anxiety_level                 0
self_esteem                   0
mental_health_history         0
depression                    0
headache                      0
blood_pressure                0
sleep_quality                 0
breathing_problem             0
noise_level                   0
living_conditions             0
safety                        0
basic_needs                   0
academic_performance          0
study_load                    0
teacher_student_relationship  0
future_career_concerns        0
social_support                0
peer_pressure                 0
extracurricular_activities    0
bullying                      0
stress_level                  0
dtype: int64
```

Gambar 5 Hasil Cek Missing Value

Pada gambar 5 diatas dapat disimpulkan bahwa tidak terdapat *missing value* dalam dataset. Setiap atribut memiliki jumlah nilai kosong sebesar 0, yang menunjukkan bahwa seluruh data telah terisi dengan lengkap. Oleh karena itu dataset siap untuk tahap analisi lebih lanjut tanpa perlu dilakukan *preprocessing* tambahan terkait nilai yang hilang.

Pada penelitian ini, proses normalisasi data tidak dilakukan karena metode yang akan digunkaan memiliki ketahannan terhadap keberadaan outlier serta skala fitur yang bervariasi. Normalisasi biasanya diterapkan untuk menyamarkan skala antar variabel, guna meningkatkan kinerja algoritma. Namun dalam hal ini, model yang digunakan seperti *Random Forest*, termasuk dalam kategori algoritma yang pada asumsi distribusi data dan memiliki ketahannan terhadap outlier. Model berbasisi pohon keputusan, membagi data berdasarkan nilai fitur tanpa memperhitungkan skala absolutnya, sehingga normalisasi tidak memberikan dampak signifikan terhadap performa model.

### *B. Split Data*

Setelah proses *preprocessing* selesai, tahap selanjutnya data dibagi menjadi data *training* dan juga data *testing*. Data *training* digunakan untuk melatih model agar belajar karakteristik data secara optimal. Sementara itu, data *testing* digunakan untuk menguji model agar dapat memberikan gambaran terhadap kemampuan model dalam mengeneralisasi data baru.

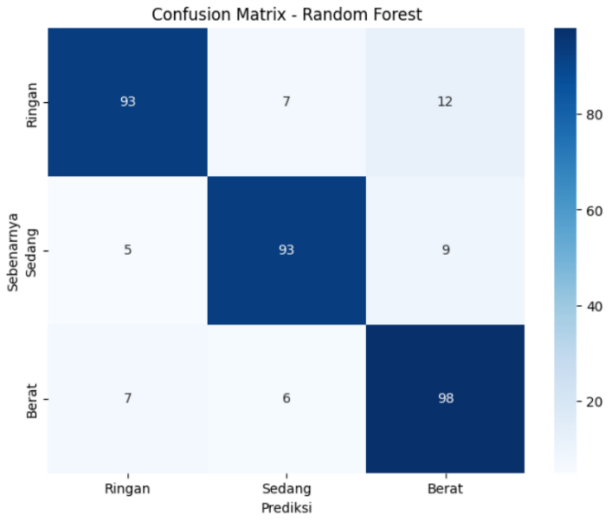
Pada penelitiana ini, pembagian dataset menggunakan rasion perbandingan 70:30, dimana 70% untuk data *training* dan 30% untuk data *testing*. Sehingga total dataset yang terdiri 1100 data, data yang dipilih secara acak untuk proses pelatihan rasio dengan 70% yaitu sebesar 770 data, sementara 30% sebesar 330 data, yang digunakan untuk menguji model yang telah dilatih. Tahap ini merupakan tapat penting sebelum dilakukan proses analisis.

### *C. Pengujian dataset dengan model*

Setelah tahap pemisahan data, algoritma *Random Forest* yang digunakan dalam penelitian ini dioptimalkan dengan mengginakan pendekatan seleksi fitur dengan metode *Backward Elimination*. Pendekatan ini digunakan untuk menikatkan kinerja model dengan memastikan bahwa hanya fitur-fitur yang relevan yang di gunakan. Seleksi fitur digunakan untuk memastikan bahwa hanya karakteristik yang relevan yang digunakan selama pelatihan, yang harusnya meningkatkan akinerja algoritma *Random Forest*.

1) Hasil Pemodelan Algoritma *Random Forest*

Pada tahap pertama penerapan klasifikasi pada algoritma *Random Forest*, Algoritma ini dipilih dalam penelitian karena memiliki keunggulan dalam menangani data yang *overfitting*. Pengujian model dilakukan mengguakan *confusion matrix*, *confusion matrix* digunakan untuk menilai kinerja model dan menghasilkan sejumlah matriks evaluasi utama, termasuk *recall*, *accuracy*, *precision* dan *f1-score*, yang memberikan pemahaman tentang seberapa baik model meprediksi setiap kelas. Hasil *confusion matrix* dari algoritma *Random Fores* ditampilkan pada gambar 6 sebagai berikut.




Gambar 6 Confusion Matrix Algoritma Random Forest

Gambar 6, menampilkan *confusion matrix* hasil klasifikasi model yang terdiri dari tiga kelas, yaitu tingkat stres ringan, tingkat stres sedang dan tingkat stress berat. Model ini memiliki akurasi 86.1%. nilai diagonal pada matris menunjukkan prediksi yang benar, yaitu 93 yang diprediksi benar sebagi ringan, 93 yang diprediksi benar sebagai sedang, dan 98 yang diprediksi benar sebagai berat. Sementara itu kesalahan prediksi terjadi pada beberapa kasus, seperti 12 kasus ringan yang diprediksi berat serta 9 kasus sedang yang diprediksi sebagau berat. Berikut pada tabel 3, ditampilkan performa model dalam mengklasifikasikan data.

TABEL 3
HASIL AKURASI RANDOM FOREST

| Pengujian pada model | Nilai | | | |
|---|---|---|---|---|
| | Precision | Recall | F1-score | Accuracy |
| Random Forest | 86.2% | 86.1% | 86.1% | 86.1% |

Tabel 3, dapat disimpulkan hasil evaluasi model *Random Forest* dalam mengklasifikasikan tingkat stres akademik, Akurasi yang di peroleh pada model *Random Forest* dengan nilai precision 86.2%, recall 86.1%, F1-score 86.1% dan accuracy 86.1%.

2) Fitur seleksi dengam *Backward Elimination* pada *Random Forest*

Pemilihan fitur menggunakan metode *Backward Elimination* digunakan untuk pemilihan fitur yang relevan dengan target dengan tujuan mengoptimalkan kinerja algoritma *Random Forest*, Langkah ini dilakukan setelah dataset di uji dengan model *Random Forest*.

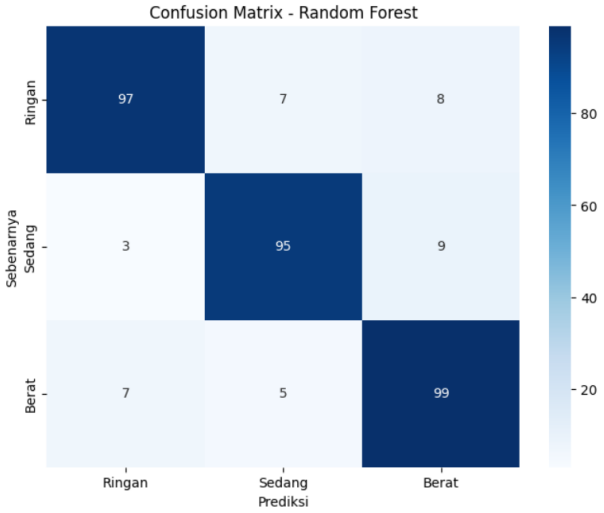
Pada tahap seleksi fitur ada 20 atribut yang akan digunakan, tahap yang pertama kita lakukan dengan menguji parametrik dengan Regresi Linier pada atribut untuk mencari hubungan antara atribut. Kemudian dilakukan uji t guna mengecek apakah setiap variable berpengaruh signifikan terhadap variabel Y. Menentukan nilai p-value sebesar 0.05 yang mana hanya fitur dengan p-value yang > 0.05 yang akan dipertahankan dalam model. Fitur dengan p-value < 0.05 secara bertahap akan di hapus, sehingga hanya fitur yang signifikan yang tersisa dalam model akhir. Model akhir memberikan hasil yang lebih akurat dan interpretatif karena hanya menggunakan fitur yang benar-benar memiliki pengaruh terhadap tingkat stress.

TABEL 4
HASIL PEMILIHAN FITUR PADA BACKWARD ELIMINATION

| Fitur yang terpilih | Fitur yang tidak terpilih |
|---|---|
| Self estem | Anxiety level |
| Depression | Mental health history |
| Headache | Blood pressure |
| Sleep quality | Breathing problem |
| Noise level | Living conditions |
| Safety | Future career concerns |
| Basic needs | Social support |
| Academic performance | |
| Study load | |
| Teacher student relationship | |
| Peer pressure | |
| Extracurricular activities | |
| Bullying | |

Dari hasil yang diperoleh pada tabel 3, terdapat sebanyak 13 fitur yang terpilih dari 20 fitur pada dataset. Fitur-fitur yang tidak dipilih dianggap memiliki konstribusi yang kecil atau redundan dalam model. Model ini hanya mempertahankan fitur yang memberikan dampak signifikan berdasarkan analisis statistik, meskipun karakteristik seperti tingkat kecemasan, riwayat kesehatan mental dan dukungan sosial berpotensi berhubungan dengan stres akademik. Menggunakan atribut yang terpilih dapat meningkatkan efisiensi model, mengurangi resiko *overfitting* dan memastikan bahwa prediksi tetap akurat. Selain itu, fitur-fitur yang terpilih masih mencakup aspek psikologis, sosial, akademik, dan lingkungan yang relevan dengan stress akademik.

Setelah dilakukan model regresi dan penentuan nilai p-value 0.05, untuk menunjukkan fitur yang relevan. Selanjutnya membanagi data testing sebesar 30% dan data testing 70% dari data baru yang sudah di seleksi fitur. Pengujian model dilakukan mengguakan *confusion matrix*, *confusion matrix* digunakan untuk menilai kinerja model dan menghasilkan sejumlah matriks evaluasi utama, termasuk *recall*, *accuracy*, *precision* dan *f1-score*, yang memberikan pemahaman tentang seberapa baik model meprediksi setiap kelas. Hasil *confusion matrix* dari algoritma *Random Fores* dengan *Backward Elimination* ditampilkan pada gambar 7.



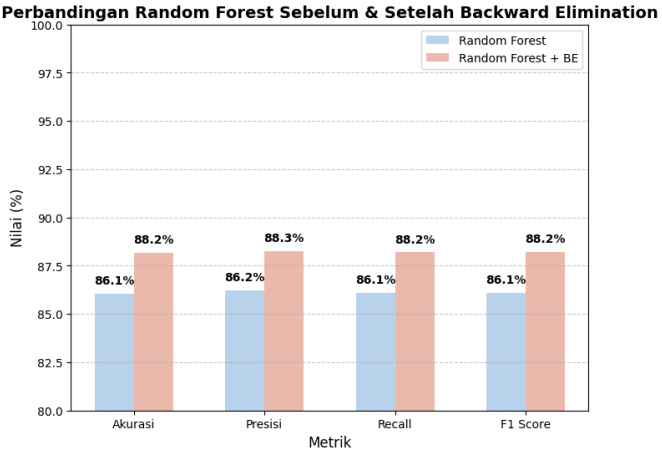

Gambar 7 Confusion matrix pada Backward Elimination + Random Forest

Gambar 7 menampilkan *confusion matrix* hasil klasifikasi model yang terdiri dari tiga kelas, yaitu tingkat stres ringan, tingkat stres sedang dan tingkat stres berat. Model ini memiliki akurasi 0.88. Nilai diagonal pada matris menunjukkan prediksi yang benar, yaitu 97 yang diprediksi benar sebagi ringan, 95 yang diprediksi benar sebagai sedang, dan 99 yang diprediksi benar sebagai berat. Sementara itu kesalahan prediksi terjadi pada beberapa kasus, seperti 8 kasus ringan yang diprediksi berat, serta 9 kasus sedang yang diprediksi sebagau berat. Berikut pada tabel 4, ditampilkan performa model dalam mengklasifikasikan data.

TABEL 5
HASIL AKURASI RANDOM FOREST + BACKWARD ELIMINATION

| Pengujian pada model | Nilai | | | |
|---|---|---|---|---|
| | Precision | Recall | F1-score | Accuracy |
| Random Forest + Backward Elimination | 88.3% | 88.2% | 88.2% | 88.2% |

Tabel 4 diatas dapat disimpulkan hasil evaluasi model dengan *Backward Elimination* pada algoritma *Random Forest*, optimasi fitur memberikan peningkatan signifikan pada kinerja model. Akurasi yang di peroleh pada model *Random Forest* dengan nilai precision 88.3%, recall 88.2%, F1-score 88.2% dan accuracy 88.2%. nilai *precision* dapat mengidentifikasi kasus positif (mahasiswa dengan gangguan stress) dengan tingkat kesalahan yang sangat rendah. Sementara nilai *recall* yang tinggi memastikan pendeteksian sebagian besar positif. Nilai *F1-score* menunjukkan keseimbangan yang hampir sempurna antara *recall* dan *precision*. Gambar 8 menampilkan hasil pemodelan pada algoritma *Random Forest* setelah dilakukan seleksi fitur menggunakan metode *Backward Elimination*.




Gambar 8 Perbandingan akurasi pengaruh seleksi fitur menggunakan Backward Elimination

Berdasarkan pada gambar 8, pada model *Random Forest* dengan Metode *Backward Elimination* memberikan hasil evaluasi yang lebih baik dibandingkan dengan metode *Random Forest* tanpa *Backward Elimination*. Model *Random Forest* tanpa *Backward Elimination* mendapatkan akurasi masing-masing sebesar 86% sedangkan pada model *Random Forest* dengam *Backward Elimination* menunjukkan hasil nilai precision 88.3%, recall 88.2%, F1-score 88.2% dan accuracy 88.2%. Hal ini menunjukkan bahwa penerapan seleksi fitur dengan *Backward Elimination* mampu meningkatkan performa model *Random Forest*. Pada model ini menunjukkan kinerja yang cukup baik dalam mengklasifikasikan tingkat stress akademik. Namun ada beberapa hal yang harus diperhatikan sebelum model ini diterapkan secara nyata. Data yang digunakan pada penelitian ini merupakan data sekunder yang diambil dari sumber dataset, dan mungkin dataset pada penelitian ini tidak sepenuhnya mewakili populasi mahasiswa dari berbagaim daerah atau universitas. Oleh sebab itu, model perlu diuji lebih lanjut dengan menggunakan dataset baru dari mahasiswa dengan berbagai latar belakang yang berbeda untuk memastikan kapasitasnya dalam menggeneralisasi.

## IV. KESIMPULAN

Berdasarkan penelitian ini, dapat disimpulkan bahwa penerapan fitur seleksi merupakan langkah untuk meningkatkan akurasi pada predikisi tingkat stres akademik. Analisis yang dilakukan dengan algoritma *Random Forest* dan *Backward Elimination* menunjukkan bahwa kedua model efektif dalam melakukan prediksi tingkat stress, pengujian model menggunakan algoritma *Random Forest* menunjukkan akurasi sebesar 86%, meskipun akurasi model ini cukup tinggi, masih terdapat fitur-fitur yang tidak relevan. Pada penelitian ini menggunakan metode *Backward Elimination* untuk memilih fitur yang relevan, guna meningkatkan akurasi dan efisiensi pada model *Random Forest*. Hasil menunjukkan peningkatkan signifikan dalam performa model *Random Forest* setelah diterapkan fitur seleksi dengan metode *Backward Elimination*, akurasi model meningkat dari 86% menjadi 88%. ini menunjukkan bahwa pemilihan fitur yang tepat dapat meningkatkan kinerja algoritma dalam klasifikasi.

## DAFTAR PUSTAKA

[1] S. Samarpita, R. Satpathy, P. K. Mishra, and A. N. Panda, "Mental Stress Classification from Brain Signals using MLP Classifier," *EAI Endorsed Trans. Pervasive Heal. Technol.*, vol. 9, no. 1, pp. 1–6, 2023, doi: 10.4108/eetpht.9.4341.

[2] K.Kesehatan, "Apa Itu Stres: Gejala, Penyebab, Pencegahan dan Pengobatan," 2024. https://ayosehat.kemkes.go.id/apa-itu-stres

[3] P. D. Ambarwati, S. S. Pinilih, and R. T. Astuti, "Gambaran Tingkat Stres Mahasiswa," *J. Keperawatan Jiwa*, vol. 5, no. 1, p. 40, 2019, doi: 10.26714/jkj.5.1.2017.40-47.

[4] D. D. W. Sari and W. Marsisno, "Klasifikasi Tingkat Stres Akademik dan Gambaran Mekanisme Koping Mahasiswa," *Semin. Nas. Off. Stat.*, vol. 2023, no. 1, pp. 203–212, 2023, doi: 10.34123/semnasoffstat.v2023i1.1691.

[5] H. B and R. Hamzah, "Faktor-Faktor Yang Berhubungan Dengan Tingkat Stres Akademik Pada Mahasiswa Stikes Graha Medika," *Indones. J. Heal. Sci.*, vol. 4, no. 2, p. 59, 2020, doi: 10.24269/ijhs.v4i2.2641.

[6] Afif Januar Ginata, Ratna Dewi Indi Astuti, and Julia Hartati, "Tingkat Stres Berdasarkan Jenis Stresor Pada Mahasiswa Tingkat Akhir Tahap Akademik Fakultas Kedokteran Unisba," *J. Ris. Kedokt.*, pp. 25–30, 2023, doi: 10.29313/jrk.vi.1915.

[7] P. H. Khrismadani, N. K. A. Sawitri, and P. O. Y. Nurhesti, "Gambaran Tingkat Stres Mahasiswa Keperawatan Universitas Udayana Dalam Proses Pembelajaran Selama Pandemi Covid-19," *Coping Community Publ. Nurs.*, vol. 10, no. 2, p. 166, 2022, doi: 10.24843/coping.2022.v10.i02.p07.

[8] W. Gamayanti, M. Mahardianisa, and I. Syafei, "Self Disclosure dan Tingkat Stres pada Mahasiswa yang sedang Mengerjakan Skripsi," *Psympathic J. Ilm. Psikol.*, vol. 5, no. 1, pp. 115–130, 2018, doi: 10.15575/psy.v5i1.2282.

[9] A. I. Noor Mahmudianti, Muhammad Riduansyah, "Journal of Health," vol. 10, no. 1, pp. 47–54, 2024.

[10] M. A. Barata *et al.*, "Perancangan Sistem Electronic Nose Berbasis," pp. 117–126, 2016.

[11] E. R. B. Sebayang, Y. H. Chrisnanto, and Melina, "Klasifikasi Data Kesehatan Mental di Industri Teknologi Menggunakan Algoritma Random Forest," *IJESPG J.*, vol. 1, no. 3, pp. 237–253, 2023.

[12] E. S. Mohamed, T. A. Naqishbandi, S. A. C. Bukhari, I. Rauf, V. Sawrikar, and A. Hussain, "A hybrid mental health prediction model using Support Vector Machine, Multilayer Perceptron, and Random Forest algorithms," *Healthc. Anal.*, vol. 3, no. July 2022, p. 100185, 2023, doi: 10.1016/j.health.2023.100185.

[13] M. R. Fanani, "Algoritma Naïve Bayes Berbasis Forward Selection Untuk Prediksi Bimbingan Konseling Siswa," *J. DISPROTEK*, vol. 11, no. 1, pp. 13–22, 2020, doi: 10.34001/jdpt.v11i1.952.

[14] C. Cahyaningtyas, D. Manongga, and I. Sembiring, "Algorithm Comparison and Feature Selection for Classification of Broiler Chicken Harvest," *J. Tek. Inform.*, vol. 3, no. 6, pp. 1717–1727, 2022, doi: 10.20884/1.jutif.2022.3.6.493.

[15] Chhabii, "Student Stress Factors: A Comprehensive Analysis," *kaggle.com*, 2022. https://www.kaggle.com/datasets/rxnach/student-stress-factors-a-comprehensive-analysis/data

[16] N. N. Sholihah and A. Hermawan, "Implementation of Random Forest and Smote Methods for Economic Status Classification in Cirebon City," *J. Tek. Inform.*, vol. 4, no. 6, pp. 1387–1397, 2023, doi: 10.52436/1.jutif.2023.4.6.1135.

[17] D. Papakyriakou and I. S. Barbounakis, "Data Mining Methods: A Review," *Int. J. Comput. Appl.*, vol. 183, no. 48, pp. 5–19, 2022, doi: 10.5120/ijca2022921884.

[18] Y. APRILLIA, "Implementasi Algoritma Naive Bayes Dengan Feature Selection Backward Elimination Dalam Pengklasifikasian Status Penderita Stunting Pada Balita," vol. 4, pp. 1–6, 2023.

[19] H. Nugroho, G. E. Yuliastuti, and A. Firman, "Klasifikasi Diagnosis Diabetes Melitus Menggunakan Metode Naïve Bayes Dengan Seleksi Fitur Backward Elimination," *J. Ilm. NERO*, vol. 8, no. 2, p. 2023, 2023.