

CESS

(Journal of Computer Engineering, System and Science)

Available online: <https://jurnal.unimed.ac.id/2012/index.php/cess>

ISSN: 2502-714x (Print) | ISSN: 2502-7131 (Online)



Klasifikasi Tingkat Stres pada Mahasiswa Tingkat Akhir menggunakan Metode Naïve Bayes di Universitas Pelita Bangsa

Stress Level Classification of Final-Year Students Using Naïve Bayes at Universitas Pelita Bangsa

Rima Puji Lestari^{1*}, Nurhadi Surojudin², Eko Budiarto³

^{1,2,3} Teknik Informatika, Universitas Pelita Bangsa

Jl. Inspeksi Kalimalang No.9, Cibatu, Cikarang Selatan, Kab Bekasi, Jawa Barat

Email: ¹rimapujilestari2@gmail.com, ²nurhadi@pelitabangsa.ac.id, ³ekobudiarto@pelitabangsa.ac.id

*Corresponding Author

ABSTRAK

Stres sering dialami oleh mahasiswa, khususnya mahasiswa tingkat akhir yang sedang menyelesaikan skripsi. Penelitian ini dilakukan untuk mengklasifikasikan tingkat stres yang sedang dihadapi oleh mahasiswa tingkat akhir Prodi Teknik Informatika angkatan 2021 di Universitas Pelita Bangsa dengan menggunakan metode *Naïve Bayes*. Berdasarkan hasil kuesioner, terdapat 300 responden berpartisipasi. Dengan menggunakan pendekatan SEMMA (*Sample, Explore, Modify, Model, Assess*), data dari 300 responden dilakukan proses *handling* sehingga tersisa 285 data yang valid, kemudian data dibagi menjadi data latih dan data uji dengan perbandingan 80:20. Tingkat stres dikategorikan menjadi 3 yaitu rendah, sedang, dan tinggi. Performa model menunjukkan bahwa model memiliki akurasi sebesar 93%, dengan nilai *precision*, *recall*, dan *F1-score* keseluruhan masing-masing sebesar 93%. Hasil klasifikasi diperoleh bahwa tingkat stres sedang merupakan kategori terbanyak yaitu sebesar 42,1%, diikuti oleh stres tinggi 29,1%, dan stres rendah 28,8%. Temuan pada penelitian ini dapat digunakan sebagai dasar pengambilan keputusan untuk penyediaan layanan dukungan psikologis dan akademik bagi mahasiswa tingkat akhir.

Kata Kunci: *Klasifikasi; Tingkat Stres; Mahasiswa Akhir; Naïve Bayes; SEMMA*

ABSTRACT

Stress is commonly experienced by students, especially final-year students who are working on their thesis. This study aims to classify the stress levels experienced by final-year students of the 2021 class in the informatics Engineering Study Program at Universitas Pelita Bangsa using the *Naïve Bayes* method. Based on the questionnaire result, a total of 300 respondents participated. Using the SEMMA approach (*Sample, Explore, Modify, Model, Assess*), data from the 300 respondents underwent a handling process, resulting in 285 valid data entries. The



data were then split into training and testing sets with an 80:20 ratio. Stress levels were categorized into three groups: low, moderate, and high. The model's performance showed an *accuracy* of 93%, with precision, recall, and F1-score each also reaching 93%. The classification result showed that the moderate stress level was the most dominant category at 42.1%, followed by high stress at 29.1%, and low stress at 28.8%. These findings can be used as a basis for decision-making in providing psychological and academic support services for final-year students.

Keywords: *Classification; Stress Level; Final-Year Students; Naïve Bayes; SEMMA*

1. PENDAHULUAN

Stres menjadi salah satu masalah utama yang dihadapi oleh mahasiswa di berbagai perguruan tinggi, khususnya mahasiswa tingkat akhir yang sedang menyelesaikan skripsi. Stres adalah perasaan tertekan dan ketegangan yang dialami oleh respon individu terhadap perubahan dalam situasi yang mengancam [1]. Jika stres tidak segera diatasi maka akan terjadi gangguan pada organ tubuh yang berakibat seseorang tidak dapat menjalankan pekerjaannya dengan baik. Faktor yang menyebabkan seseorang stres tidak hanya berasal dari faktor internal, tetapi juga faktor eksternal. Adapun faktor internal mencakup biologi dan psikologi, sedangkan faktor eksternal berasal dari lingkungan [2].

Mahasiswa tingkat akhir yang sedang menyelesaikan skripsi sering kali berada dalam tekanan akademik yang tinggi, hal ini dikarenakan beban tugas, tuntutan kelulusan, maupun faktor lainnya. Jika tidak dikenali sejak dini, stres yang dialami dapat berdampak negatif terhadap proses penyusunan skripsi maupun kesehatan mental mahasiswa itu sendiri. Menurut data WHO (*World Health Organization*) pada tahun 2020 diketahui jumlah kasus mahasiswa tingkat akhir mengalami stres saat menyusun tugas akhir di dunia sebesar 38% - 71%, sedangkan di Asia mencapai 39,6%. Berdasarkan data BPS (Badan Pusat Statistik) pada tahun 2019, mahasiswa tingkat akhir di Indonesia mengalami stres sebesar 61,3% [3]. Namun, saat ini diketahui belum adanya sistem yang dapat mengidentifikasi tingkat stres berdasarkan data yang ada. Oleh karena itu dilakukan klasifikasi tingkat stres mahasiswa tingkat akhir di Universitas Pelita Bangsa dengan tiga kategori tingkat stres yaitu rendah, sedang, dan tinggi menggunakan metode *Naïve Bayes*.

Penelitian sebelumnya dilakukan penelitian menggunakan metode *Naïve Bayes* terhadap 23 responden melalui kuesioner yang disebarakan berbasis Google Form, diperoleh hasil bahwa sebagian besar mahasiswa memiliki gejala gangguan mental tingkat sedang saat menjelang skripsi, dengan prediksi tingkat stres ringan 30%, sedang 56,25%, dan berat 13,75%, serta mencapai akurasi model sebesar 87,5% [4]. Penelitian lain menyebutkan dengan menggunakan sumber data hasil wawancara dan kuesioner kepada 34 responden, diperoleh hasil akurasi data dari model *Naïve Bayes* sebanyak 85%, dengan diagnosa 29 benar dari 34 data responden [5].

Pada penelitian ini dilakukan klasifikasi tingkat stres mahasiswa tingkat akhir dari Prodi Teknik Informatika angkatan 2021 di Universitas Pelita Bangsa menggunakan metode *Naïve Bayes*. Sumber data diambil melalui kuesioner yang disebarakan berbasis Google Form dengan penggunaan Skala Likert. Data tersebut diolah menggunakan bahasa pemrograman Python di Google Colab dengan pendekatan SEMMA. Evaluasi model dilakukan menggunakan akurasi, *confusion matrix*, dan visualisasi data untuk menilai kinerja klasifikasi berdasarkan data secara terukur.

Perbedaan penelitian ini dengan penelitian sebelumnya yang juga menggunakan metode *Naïve Bayes* dan kuesioner yaitu pada penerapan pendekatan SEMMA yang digunakan untuk mengelola dan menganalisis dengan jumlah data responden lebih banyak dibandingkan dengan penelitian sebelumnya, dilakukan konversi skor dari Skala Likert untuk menentukan kategori tingkat stres, serta menambahkan visualisasi data dari hasil yang diperoleh.

Dengan dilakukan klasifikasi tingkat stres terhadap mahasiswa tingkat akhir di Universitas Pelita Bangsa menggunakan metode *Naïve Bayes* dengan pendekatan SEMMA, penelitian ini bertujuan untuk mengembangkan model klasifikasi yang sistematis dan terukur. Hasil penelitian ini diharapkan dapat memberikan gambaran terkait kondisi mahasiswa, sehingga pihak universitas dapat mengambil langkah yang tepat dalam memberikan dukungan akademik maupun psikologis sesuai dengan tingkat stres yang dialami mahasiswa.

2. TINJAUAN TEORI

2.1 Klasifikasi

Klasifikasi adalah pengelompokan objek kedalam kelas tertentu berdasarkan kelompok [6]. Klasifikasi merupakan salah satu teknik yang ada pada *data mining* dengan menggunakan metode *supervised learning* yang membutuhkan data training berlabel. Beberapa teknik klasifikasi yang banyak digunakan diantaranya *neural network*, *decision tree*, *naïve bayes classifier*, *support machine*, dan *rule-based classifier* [7].

Pada klasifikasi data dilakukan proses untuk menemukan model atau fungsi yang menjelaskan dan membedakan kelas beserta konsepnya [8]. Proses klasifikasi banyak digunakan dalam berbagai bidang, diantaranya analisis sentimen, prediksi pelaku, dan pengelompokan data tekstual, hal ini berguna untuk menghasilkan informasi yang mendukung pengambilan keputusan menjadi lebih efisien [9].

2.2 Tingkat Stres

Tingkat stres dapat dikategorikan berat atau ringan nya stres yang dialami oleh seseorang [10], dalam penelitian ini khususnya mahasiswa. Banyak faktor stres yang dialami baik secara biologi maupun psikologi. Gejala stres dapat terbagi menjadi tiga aspek, pada gejala fisik dapat menyebabkan terganggu nya waktu tidur dan adanya perubahan waktu makan. Gejala emosional bisa berupa perubahan suasana hati, sehingga mudah marah, cemas, dan kurang semangatnya ber aktifitas. Adapun yang terakhir gejala tidak mampu berkonsentrasi, merasa bingung, serta meningkatnya pikiran negatif seseorang [11].

2.3 Naïve Bayes

Naïve Bayes merupakan salah satu metode dalam *machine learning* yang digunakan untuk klasifikasi dengan menerapkan probabilitas dan statistik. Metode ini ditemukan oleh ilmuwan asal Inggris yaitu Thomas Bayes. *Naïve Bayes* dianggap cukup efektif dalam klasifikasi serta memiliki tingkat akurasi yang baik dibandingkan dengan metode klasifikasi yang lain. Adapun rumus *Naïve Bayes* adalah sebagai berikut [12]:

$$P(H|X) = \frac{P(X|H) \times P(H)}{P(X)}$$

Keterangan:

$P(H|X)$ = Probabilitas H berdasarkan kondisi X

$P(X|H)$ = Probabilitas nilai X berdasarkan dari kondisi hipotesis H
 $P(H)$ = Probabilitas hipotesis H
 $P(X)$ = Probabilitas hipotesis X

Metode ini bekerja dengan berasumsi bahwa nilai-nilai atribut yang bersifat independen secara kondisional terhadap kelas target jika nilai keluaran diketahui. Hal ini menyederhanakan asumsi untuk perhitungan probabilitas setiap fitur sehingga menjadi lebih efisien untuk diterapkan pada dataset [13].

2.4 SEMMA (Sample, Explore, Modify, Model, Assess)

Pendekatan pemrosesan data yang digunakan dalam penelitian ini menggunakan SEMMA (*Sample, Explore, Modify, Model, Assess*). SEMMA merupakan metode data mining yang mudah dipahami dan diterapkan sebagai referensi sebuah proyek. Proses ini dimulai dari mengumpulkan data dan informasi (*Sample*), tahap mencari kumpulan data terkait ide yang akan dibangun (*Explore*), proses modifikasi data dan pengelompokan variable (*Modify*), proses memodelkan data (*Model*), dan proses evaluasi data (*Assess*) [14].

2.5 Confusion Matrix

Confusion Matrix adalah sebuah tabel yang menjelaskan klasifikasi jumlah data uji yang benar dan data uji yang salah. Contoh *Confusion Matrix* dapat dilihat pada Tabel 1 [15].

Tabel 1. <i>Confusion Matrix</i>			
		Kelas Prediksi	
		1	0
Kelas	1	TP	FN
Sebenarnya	0	FP	TN

Keterangan:

TP (*True Positive*) = Prediksi benar untuk kelas 1
TN (*True Negative*) = Prediksi benar untuk kelas 0
FP (*False Positive*) = Prediksi salah untuk kelas 1
FN (*False Negative*) = Prediksi salah untuk kelas 0

Adapun rumus *Confusion Matrix* yang digunakan untuk menghitung nilai *accuracy*, *precision*, dan *recall* adalah berikut:

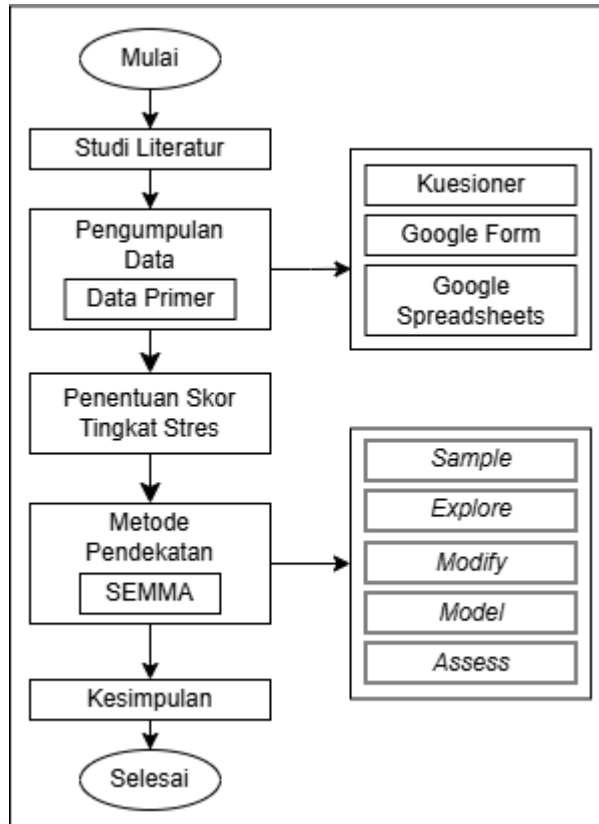
$$Accuracy = \frac{TP+TN}{Total}$$

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

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

3. METODE PENELITIAN

Pada penelitian ini menggunakan *Naïve Bayes* untuk melakukan klasifikasi tingkat stres mahasiswa tingkat akhir dan mendapatkan akurasi dari pemodelan data. Proses yang dilakukan dapat dilihat pada Gambar 1.



Gambar 1. Flowchart Alur Penelitian Klasifikasi Tingkat Stres Mahasiswa Tingkat Akhir

- 1) Studi literatur pada penelitian ini dilakukan dengan mencari data pendukung yang relevan dan referensi yang terpercaya agar penelitian lebih fokus dan terarah.
- 2) Tahap pengumpulan data dilakukan dengan menggunakan data primer, yaitu data yang didapat oleh peneliti secara langsung berdasarkan sumber datanya [16]. Kuesioner disebarakan kepada responden berbasis Google Form dengan pertanyaan Skala Likert yang memiliki rentang skor 1 sampai 5, kemudian jawaban tersebut akan tersimpan secara otomatis di Google Spreadsheets.
- 3) Penentuan skor tingkat stres dilakukan dengan menghitung total minimal, total maksimal, *range*, dan menghitung nilai panjang kelas interval pada kuesioner.
- 4) Dengan menggunakan metode pendekatan SEMMA, tahap pengolahan data dan evaluasi model dilakukan secara berurutan dimulai dari tahap *Sample*, *Explore*, *Modify*, *Model*, dan *Assess*. Metode *Naïve Bayes* dipilih karena selain sederhana dan cepat, cukup efektif untuk klasifikasi data seperti kuesioner, hal ini berdasarkan pada probabilitas dari tiap fitur terhadap kelas.
- 5) Hasil klasifikasi dan nilai akurasi yang didapat akan disimpulkan untuk mengetahui sejauh mana tingkat stres yang dimiliki oleh responden.

4. HASIL DAN PEMBAHASAN

4.1 Pengumpulan Data

Pengumpulan data dilakukan dengan cara menyebarkan kuesioner kepada responden berbasis Google Form. Jawaban dari responden akan langsung tersimpan di Google Spreadsheets. Peneliti mengajukan 10 pertanyaan berbentuk Skala Likert dengan rentang skor 1 sampai 5. Adapun pertanyaan yang terdapat dalam kuesioner tercantum pada Tabel 2.

Tabel 2. Pertanyaan Kuesioner

No	Pertanyaan
1	Seberapa sering Anda merasa beban skripsi terlalu berat?
2	Seberapa sulit Anda memahami bimbingan dari dosen pembimbing?
3	Seberapa sering Anda merasa sulit mengatur waktu antara skripsi dan aktivitas lain?
4	Seberapa sering Anda menunda-nunda pengerjaan skripsi?
5	Seberapa puas Anda dengan dukungan dari teman atau keluarga terkait skripsi?
6	Apakah Anda merasa mendapat dukungan yang cukup dari dosen pembimbing?
7	Dalam sebulan terakhir, seberapa sering Anda merasa kewalahan memikirkan skripsi?
8	Dalam sebulan terakhir, seberapa sering Anda sulit tidur karena memikirkan skripsi?
9	Dalam sebulan terakhir, apakah Anda merasa lelah secara emosional karena skripsi?
10	Seberapa sering Anda merasa cemas tentang masa depan setelah lulus?

Total populasi mahasiswa tingkat akhir Prodi Teknik Informatika angkatan 2021 di Universitas Pelita Bangsa berjumlah 328 mahasiswa. Pada penelitian ini terdapat 300 mahasiswa berpartisipasi menjadi responden. Data tersebut diperoleh dari tanggal 10 Desember 2024 sampai 21 Mei 2025, seperti pada Gambar 2.

Timestamp	Email	Jenis Kelamin	Usia	Semester	1. Seberapa sering Anda merasa beban skripsi terlalu berat?	2. Seberapa sulit Anda memahami bimbingan dari dosen pembimbing?	3. Seberapa sering Anda merasa sulit mengatur waktu antara skripsi dan aktivitas lain?	4. Seberapa sering Anda menunda-nunda pengerjaan skripsi?	5. Seberapa puas Anda dengan dukungan dari teman atau keluarga terkait skripsi?	6. Apakah Anda merasa mendapat dukungan yang cukup dari dosen pembimbing?	7. Dalam sebulan terakhir, seberapa sering Anda merasa kewalahan memikirkan skripsi?	8. Dalam sebulan terakhir, seberapa sering Anda sulit tidur karena memikirkan skripsi?	9. Dalam sebulan terakhir, apakah Anda merasa lelah secara emosional karena skripsi?	10. Seberapa sering Anda merasa cemas tentang masa depan setelah lulus?
0 10/12/2024 16:22:23	aldira05@gmail.com	Laki - Laki	> 20 Tahun	Semester 7	3	2	5	4	3	4	4	2	2	5
1 10/12/2024 16:30:31	auramutiarani7@gmail.com	Perempuan	> 20 Tahun	Semester 7	3	3	5	4	4	3	5	5	5	5
2 10/12/2024 16:33:43	pringgandanirizal37@gmail.com	Laki - Laki	> 20 Tahun	Semester 7	1	1	5	3	5	5	2	4	1	5
3 10/12/2024 16:45:13	sidikrifi481@gmail.com	Laki - Laki	> 20 Tahun	Semester 7	3	3	2	2	2	4	3	3	4	5
4 10/12/2024 16:49:22	railhanherfiansyah22@gmail.com	Laki - Laki	> 20 Tahun	Semester 7	4	3	5	5	5	3	3	2	4	5

Gambar 2. Hasil Pengumpulan Data

4.2 Penentuan Skor Tingkat Stres

Penentuan skor tingkat stres dilakukan dengan menghitung total minimal, total maksimal, *range*, dan menghitung nilai panjang kelas interval. Dengan rentang skor Skala Likert 1 sampai 5, selanjutnya akan dilakukan pengkategorian tingkat stres. Seperti penjelasan berikut ini [17]:

- 1) Menghitung total minimal

Pada penelitian ini terdapat skor minimal yaitu 1, serta total pertanyaan pada kuesioner sebanyak 10. Adapun untuk memperoleh skor minimal dilakukan perhitungan:

$$\begin{aligned} \text{Total Min} &= \text{Skor Minimal Skala Likert} \times \text{Total Pertanyaan} \\ \text{Total min} &= 1 \times 10 = 10 \end{aligned}$$

- 2) Menghitung total maksimal

Pada penelitian ini terdapat skor maksimal yaitu 5. Adapun untuk memperoleh skor maksimal dilakukan perhitungan:

$$\begin{aligned} \text{Total Maksimal} &= \text{skor maksimal skala likert} * \text{total pertanyaan} \\ \text{Total Maksimal} &= 5 \times 10 = 50 \end{aligned}$$

- 3) Menentukan rentang (*range*)

Untuk menentukan rentang, maka dilakukan perhitungan:

$$\begin{aligned} \text{Range} &= \text{Total Maksimal} - \text{Total Minimal} \\ \text{Range} &= 50 - 10 = 40 \end{aligned}$$

- 4) Menghitung nilai panjang kelas interval (*p*)

Untuk menghitung nilai interval, maka dilakukan perhitungan:

$$p = \frac{40}{3} \quad p = 13,33$$

- 5) Melakukan pengkategorian

Pada penelitian ini terdapat 3 kategori tingkat stres yaitu rendah, sedang, dan tinggi. Oleh karena itu, dengan panjang kelas interval (*p*) = 13,33 dan dengan dimulai batas bawah 10, maka distribusi tingkat stres dapat dilihat pada Tabel 3.

Tabel 3. Distribusi Tingkat Stres

Proses	<i>p</i>	Kategori
$10 + 13,33 = 23,33 = 23$	$10 - 23$	Rendah
$23,33 + 13,33 = 36,66 = 37$	$24 - 37$	Sedang
$36,66 + 13,33 = 49,99 = 50$	$38 - 50$	Tinggi

4.3 Pengolahan Data dan Evaluasi Model

Pada tahap pengolahan data dan evaluasi model dilakukan metode pendekatan SEMMA (*Sample, Explore, Modify, Model, Assess*). Berikut merupakan tahapan pada penelitian ini menggunakan metode pendekatan SEMMA:

1) *Sample*

Dataset pada penelitian ini diperoleh dari hasil kuesioner berbasis Google Form yang tersimpan di Google Spreadsheets. Kemudian data diambil menggunakan teknik *data fetching* dengan *library gspread* pada bahasa pemrograman python di Google Colab. Hal ini dilakukan untuk integrasi Google Spreadsheets secara langsung dengan Google Colab, agar dataset menjadi format CSV. Pada proses ini juga dilakukan penghapusan kolom atau baris (*drop*) dari dataset yang tidak diperlukan. Seperti pada Gambar 3.

	1. Seberapa sering Anda merasa beban skripsi terlalu berat?	2. Seberapa sulit Anda memahami bimbingan dari dosen pembimbing?	3. Seberapa sering Anda merasa sulit mengatur waktu antara skripsi dan aktivitas lain?	4. Seberapa sering Anda menunda-nunda pengerjaan skripsi?	5. Seberapa puas Anda dengan dukungan dari teman atau keluarga terkait skripsi?	6. Apakah Anda merasa mendapat dukungan yang cukup dari dosen pembimbing?	7. Dalam sebulan terakhir, seberapa sering Anda merasa kewalahan memikirkan skripsi?	8. Dalam sebulan terakhir, seberapa sering Anda sulit tidur karena memikirkan skripsi?	9. Dalam sebulan terakhir, apakah Anda merasa lelah secara emosional karena skripsi?	10. Seberapa sering Anda merasa cemas tentang masa depan setelah lulus?
0	3	2	5	4	3	4	4	2	2	5
1	3	3	5	4	4	3	5	5	5	5
2	1	1	5	3	5	5	2	4	1	5
3	3	3	2	2	2	4	3	3	4	5
4	4	3	5	5	5	3	3	2	4	5
...	...	...	...	...	...	...	...	...	...	...
295	5	5	5	5	5	5	5	5	5	5
296	2	1	2	1	3	1	2	2	2	1
297	5	5	5	5	5	5	5	5	5	5
298	4	5	3	4	5	4	4	5	3	5
299	3	4	3	4	3	4	4	3	4	4

300 rows x 10 columns

Gambar 3. *Drop Dataset*

Selanjutnya mengubah kolom pertanyaan dari 1 sampai 10 menjadi q1 sampai q10. Hal ini dilakukan supaya mempermudah penulisan kode serta mudah digunakan untuk visualisasi dan model. Seperti pada Gambar 4.

	q1	q2	q3	q4	q5	q6	q7	q8	q9	q10
0	3	2	5	4	3	4	4	2	2	5
1	3	3	5	4	4	3	5	5	5	5
2	1	1	5	3	5	5	2	4	1	5
3	3	3	2	2	2	4	3	3	4	5
4	4	3	5	5	5	3	3	2	4	5
...	...	...	...	...	...	...	...	...	...	...
295	5	5	5	5	5	5	5	5	5	5
296	2	1	2	1	3	1	2	2	2	1
297	5	5	5	5	5	5	5	5	5	5
298	4	5	3	4	5	4	4	5	3	5
299	3	4	3	4	3	4	4	3	4	4

300 rows x 10 columns

Gambar 4. Mengubah Kolom Pertanyaan

2) *Explore*

Pada tahap *explore* dilakukan pengecekan *missing value* untuk mengetahui adanya data yang hilang. Selain itu, terdapat data duplikat pada dataset ini sebanyak 15 data, seperti pada Gambar 5.

	q1	q2	q3	q4	q5	q6	q7	q8	q9	q10
163	3	3	3	3	3	3	3	3	3	3
166	5	5	5	5	5	5	5	5	5	5
168	4	4	4	4	4	4	4	4	4	4
196	2	2	2	2	2	2	2	2	2	5
206	5	5	5	5	5	5	5	5	5	5
215	3	3	3	3	3	3	3	3	3	3
231	4	4	4	4	4	4	4	4	4	4
232	5	5	5	5	5	5	5	5	5	5
234	5	5	5	5	5	5	5	5	5	5
250	2	2	2	2	2	2	2	2	2	2
254	3	3	3	3	3	3	3	3	3	3
277	4	4	4	4	4	4	4	4	4	4
291	3	3	3	3	3	3	3	3	3	3
295	5	5	5	5	5	5	5	5	5	5
297	5	5	5	5	5	5	5	5	5	5

Gambar 5. Data Duplikat

3) Modify

Pada tahap *modify* dilakukan *handling* data duplikat, hal ini dikarenakan data duplikat dapat menyebabkan bias serta mengganggu akurasi model. Pada Gambar 6 terlihat bahwa setelah melakukan *handling* data duplikat, tersisa 285 data dari 300 data sebenarnya.

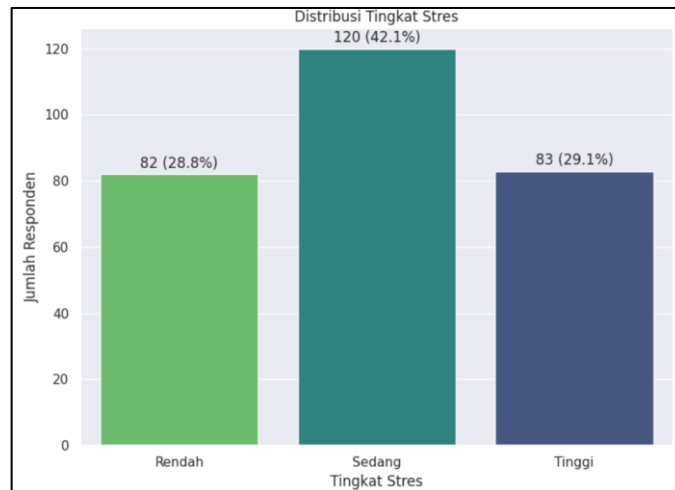
	q1	q2	q3	q4	q5	q6	q7	q8	q9	q10
0	3	2	5	4	3	4	4	2	2	5
1	3	3	5	4	4	3	5	5	5	5
2	1	1	5	3	5	5	2	4	1	5
3	3	3	2	2	2	4	3	3	4	5
4	4	3	5	5	5	3	3	2	4	5
...	...	...	...	...	...	...	...	...	...	...
280	3	2	3	4	2	2	3	4	2	5
281	1	1	2	5	5	5	5	5	5	3
282	2	1	2	1	3	1	2	2	2	1
283	4	5	3	4	5	4	4	5	3	5
284	3	4	3	4	3	4	4	3	4	4

285 rows × 10 columns

Gambar 6. Handling Data Duplikat

Selanjutnya dilakukan proses *labelling* untuk menentukan tingkat stres menggunakan fungsi klasifikasi dengan rentang skor yang telah ditentukan, kemudian data akan dikelompokkan kedalam kategori rendah, sedang, dan tinggi. Hasil menunjukkan bahwa kategori tingkat stres sedang memiliki jumlah terbanyak yaitu 120 orang dengan

persentase 42,1%, stres tinggi 83 orang dengan persentase 29,1%, dan stres rendah 82 orang dengan persentase 28,8%. Dapat dilihat pada Gambar 7.



Gambar 7. Visualisasi Tingkat Stres Responden

Dominasi tingkat stres sedang pada penelitian ini mencerminkan tekanan akademik yang cukup tinggi namun masih dapat dihadapi oleh sebagian besar mahasiswa tingkat akhir. Hasil ini sejalan dengan studi sebelumnya yang menunjukkan distribusi stres ringan 30%, sedang 56,25%, dan berat 13,75%. Dibandingkan dengan studi tersebut, penelitian ini menyajikan hasil dalam bentuk visualisasi, sehingga menghasilkan analisis yang mudah dipahami secara sistematis.

Setelah itu dilakukan proses *one hot encoding* untuk mengubah variabel kategorikal menjadi format numerik biner, hal ini dilakukan agar data dapat digunakan oleh algoritma *machine learning*, khususnya pada penelitian ini *Naïve Bayes*. Dengan kategori numerik, rendah = 1, sedang = 2, dan tinggi = 3. Hasil dapat dilihat pada Gambar 8.

	q1	q2	q3	q4	q5	q6	q7	q8	q9	q10	total_skor	tingkat_stres
0	3	2	5	4	3	4	4	2	2	5	34	2
1	3	3	5	4	4	3	5	5	5	5	42	3
2	1	1	5	3	5	5	2	4	1	5	32	2
3	3	3	2	2	2	4	3	3	4	5	31	2
4	4	3	5	5	5	3	3	2	4	5	39	3

Gambar 8. One Hot Encoding

4) Model

Pada tahap *model*, dataset dibagi menjadi data latih (*train*) dan data uji (*test*) dengan perbandingan 80:20 menggunakan fungsi *train_test_split*. Data latih digunakan untuk melatih model *Naïve Bayes*, sedangkan data uji digunakan untuk mengukur performa model. Pada Tabel 4 terlihat perbandingan data latih dan data uji yang digunakan.

Jenis Data	Jumlah
Data Latih	228

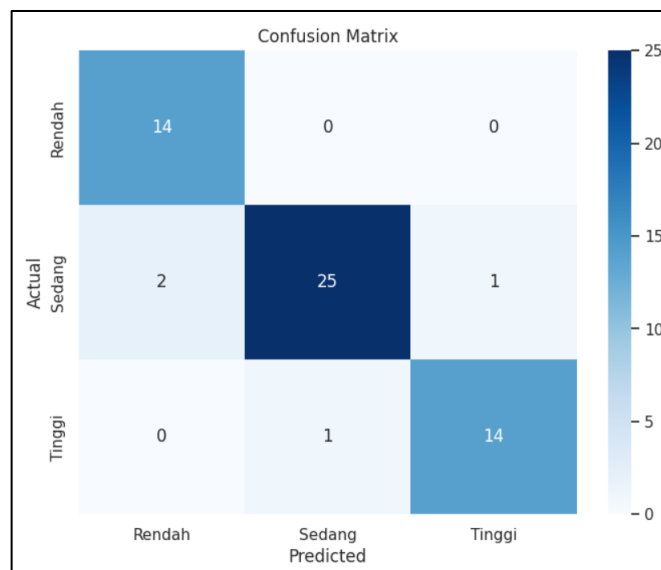
5) Assess

Pada tahap *assess* dilakukan evaluasi terhadap performa model klasifikasi yang telah dibangun. Evaluasi awal dilakukan dengan menghitung akurasi model dari perbandingan antara jumlah prediksi yang benar dengan total jumlah data uji. Berdasarkan hasil pengujian, diperoleh akurasi sebesar 93%, hal ini menunjukkan bahwa model memiliki performa yang sangat baik dalam mengklasifikasikan tingkat stres mahasiswa. Seperti pada Gambar 9.

Akurasi Model: 0.93				
Laporan Klasifikasi:				
	precision	recall	f1-score	support
1	0.88	1.00	0.93	14
2	0.96	0.89	0.93	28
3	0.93	0.93	0.93	15
accuracy			0.93	57
macro avg	0.92	0.94	0.93	57
weighted avg	0.93	0.93	0.93	57

Gambar 9. Akurasi Model *Naïve Bayes*

Evaluasi selanjutnya menggunakan *confusion matrix* untuk melihat performa klasifikasi model terhadap masing-masing kategori tingkat stres yaitu rendah, sedang, dan tinggi. Pada matriks ini digambarkan jumlah data yang berhasil diklasifikasikan dengan benar, maupun jumlah data yang salah diklasifikasikan. *Confusion matrix* dapat dilihat pada Gambar 10.



Gambar 10. *Confusion Matrix*

Berdasarkan *confusion matrix* tersebut, model berhasil mengklasifikasikan 14 data stres rendah dengan benar, pada data stres sedang model berhasil mengklasifikasikan 25 data benar dari 28 data, serta untuk data stres tinggi model berhasil mengklasifikasikan 14 data benar dari 15 data.

5. KESIMPULAN

Pada penelitian ini menunjukkan bahwa evaluasi model *Naïve Bayes* mampu melakukan klasifikasi dengan sangat baik dan seimbang sehingga menghasilkan akurasi sebesar 93%, dengan nilai *precision*, *recall*, dan *F1-score* secara keseluruhan masing-masing sebesar 93%. Klasifikasi tingkat stres mahasiswa tingkat akhir Prodi Teknik Informatika di Universitas Pelita Bangsa melalui kuesioner yang disebarakan berbasis Google Form dengan Skala Likert 1 sampai 5, diperoleh 300 data dan sudah dilakukan proses *handling* menjadi 285 data, menghasilkan data tingkat stres sedang sebanyak 120 responden dengan persentase 42,1%, stres tinggi sebanyak 83 responden dengan persentase 29,1%, dan stres rendah sebanyak 82 responden dengan persentase 28,8%. Hasil klasifikasi menunjukkan pentingnya peran universitas dalam memberikan dukungan yang sesuai untuk mahasiswa tingkat akhir. Diantaranya dapat menyediakan layanan konseling atau bimbingan akademik yang terstruktur agar mahasiswa tingkat akhir dapat menghadapi tekanan dengan lebih baik selama proses penyusunan skripsi. Metode ini juga dapat diterapkan menjadi bagian dari sistem pendukung keputusan di bidang pendidikan dan psikologi untuk memantau kondisi mental mahasiswa secara objektif.

DAFTAR PUSTAKA

- [1] L. N. Hidayati dan M. Harsono, "Tinjauan Literatur Mengenai Stres Dalam Organisasi," *J. ILMU Manaj.*, vol. 18, no. 1, Art. no. 1, 2021, doi: 10.21831/jim.v18i1.39339.
- [2] W. Kurniawati, "Manajemen Stress Era New Normal Dalam Dunia Pendidikan," *J. MUBTADIIN*, vol. 8, no. 01, Art. no. 01, Mei 2022. Tersedia pada: <https://journal.an-nur.ac.id/index.php/mubtadiin/article/view/175>
- [3] V. N. Wahyuni dan N. Handayani, "Hubungan kepercayaan diri dengan stres mahasiswa tingkat akhir keperawatan anestesiologi Universitas 'Aisyiyah Yogyakarta,'" *Pros. Semin. Nas. Penelit. Dan Pengabd. Kpd. Masy. LPPM Univ. Aisyiyah Yogyakarta.*, vol. 2, hlm. 1418–1423, Okt 2024.
- [4] N. A. Sivi, F. Dwiatmoko, D. Utami, dan A. Hafiz, "Penerapan Metode Naive Bayes Untuk Memprediksi Tingkat Kesehatan Mental Mahasiswa UNU Lampung Menjelang Skripsi," *J. Inf. Dan Komput.*, vol. 13, no. 01, hlm. 174–186, Apr 2025, doi: 10.35959/jik.v13i01.687.
- [5] W. S. Wisnugraha, I. N. Farida, dan M. A. D. Widyadara, "Implementasi Algoritma Naïve Bayes Dalam Menentukan Diagnosa Tingkat Depresi Mahasiswa Akhir Terhadap Pengerjaan Skripsi," dalam *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, 2023, hlm. 919–928. Tersedia pada: <https://proceeding.unpkediri.ac.id/index.php/inotek/article/view/3517>
- [6] H. F. Putro, R. T. Vlandari, dan W. L. Y. Saptomo, "Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan," *J. Teknol. Inf. Dan Komun. TIKomSiN*, vol. 8, no. 2, 2020. Tersedia pada: <https://p3m.sinus.ac.id/jurnal/index.php/TKomSiN/article/view/500>
- [7] P. B. N. Setio, D. R. S. Saputro, dan B. Winarno, "Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4. 5," dalam *PRISMA, Prosiding Seminar Nasional Matematika*,

- 2020, hlm. 64–71. Tersedia pada: <https://journal.unnes.ac.id/sju/prisma/article/download/37650/15478>
- [8] N. B. Putri dan A. W. Wijayanto, “Analisis Komparasi Algoritma Klasifikasi Data Mining Dalam Klasifikasi Website Phishing,” *Komputika J. Sist. Komput.*, vol. 11, no. 1, hlm. 59–66, 2022.
- [9] R. R. Adhitya, W. Witanti, dan R. Yuniarti, “Perbandingan Metode Cart Dan Naïve Bayes Untuk Klasifikasi Customer Churn,” *INFOTECH J.*, vol. 9, no. 2, hlm. 307–318, 2023.
- [10] R. D. A, F. P. Idris, F. A. Gobel, A. Asrina, dan H. Rahman, “The Faktor Yang Berhubungan Dengan Tingkat Stres Mahasiswa Dalam Menyusun Skripsi,” *Window Public Health J.*, hlm. 567–577, Agu 2023, doi: 10.33096/woph.v4i4.1115.
- [11] M. R. Rahadiansyah dan A. Chusairi, “Pengaruh Dukungan Sosial Teman Sebaya terhadap Tingkat Stres Mahasiswa yang Mengerjakan Skripsi,” *Bul. Ris. Psikol. Dan Kesehat. Ment. BRPKM*, vol. 1, no. 2, hlm. 1290–1297, Agu 2021, doi: 10.20473/brpkm.v1i2.29077.
- [12] R. F. Mahendra, G. A. Buntoro, dan A. Y. Astuti, “Implementasi Algoritma Naive Bayes Pada Sistem Pakar Diagnosa Tingkat Stres Mahasiswa Akhir,” *J. Responsif Ris. Sains Dan Inform.*, vol. 6, no. 2, hlm. 293–301, Agu 2024, doi: 10.51977/jti.v6i2.1810.
- [13] A. Mukti, A. D. Hadiyanti, A. Nurlaela, dan J. Panjaitan, “Sistem Analisa Sentiment Bakal Calon Presiden 2024 Menggunakan Metode NLP Berbasis Web,” *SOSCIED*, vol. 6, no. 1, hlm. 128–140, 2023.
- [14] Y. A. Suwitonono dan F. J. Kaunang, “Implementasi Algoritma Convolutional Neural Network (CNN) Untuk Klasifikasi Daun Dengan Metode Data Mining SEMMA Menggunakan Keras,” *J. Komtika Komputasi Dan Inform.*, vol. 6, no. 2, hlm. 109–121, 2022.
- [15] D. Normawati dan S. A. Prayogi, “Implementasi Naïve Bayes classifier dan confusion matrix pada analisis sentimen berbasis teks pada Twitter,” *J-SAKTI J. Sains Komput. Dan Inform.*, vol. 5, no. 2, hlm. 697–711, 2021.
- [16] A. R. Fadilla dan P. A. Wulandari, “Literature review analisis data kualitatif: tahap pengumpulan data,” *Mitita J. Penelit.*, vol. 1, no. 3, hlm. 34–46, 2023.
- [17] A. A. Santika, T. H. Saragih, dan M. Muliadi, “Penerapan Skala Likert pada Klasifikasi Tingkat Kepuasan Pelanggan Agen Brilink Menggunakan Random Forest,” *J. Sist. Dan Teknol. Inf. JustIN*, vol. 11, no. 3, hlm. 405, Jul 2023, doi: 10.26418/justin.v11i3.62086.