

# Prediksi Risiko Kesehatan Mental Mahasiswa Menggunakan Klasifikasi Naive Bayes

Fithra Aditya Sumantri, Yulison Herry Chrisnanto\*, Melina

Fakultas Sains dan Informatika, Program Studi Teknik Informatika, Universitas Jenderal Achmad Yani, Cimahi, Indonesia

Email: <sup>1</sup>fithraaditya21@if.unjani.ac.id, <sup>2,\*</sup>yhc@if.unjani.ac.id, <sup>3</sup>melina@lecture.unjani.ac.id

Email Penulis Korespondensi: yhc@if.unjani.ac.id

Submitted 22-05-2025; Accepted 30-06-2025; Published 30-06-2025

## Abstrak

Kesehatan mental mahasiswa menjadi perhatian yang semakin meningkat karena tekanan akademik, emosional, dan sosial turut berkontribusi terhadap peningkatan risiko psikologis. Penelitian ini bertujuan untuk mengklasifikasikan tingkat risiko kesehatan mental mahasiswa ke dalam tiga kategori: Rendah, Sedang, dan Tinggi, menggunakan algoritma klasifikasi *Naive Bayes*. *Dataset* yang digunakan terdiri dari 1.000 entri dan 11 variabel utama yang mencakup faktor akademik, psikologis, dan perilaku. Tahapan pra-pemrosesan mencakup pembersihan data, label *encoding*, normalisasi, dan pelabelan berbasis aturan untuk menentukan kelas target. Proses pelatihan dan pengujian model dilakukan dengan pembagian data stratifikasi untuk menjaga distribusi kelas. Model awal menghasilkan akurasi klasifikasi sebesar 88,67%, dengan nilai *F1-score* rata-rata makro sebesar 0,87 dan rata-rata tertimbang sebesar 0,88. Optimasi model dilakukan menggunakan *Grid Search* dengan *k-fold cross-validation*, namun tidak menunjukkan peningkatan signifikan, yang mengindikasikan bahwa model telah berada pada konfigurasi yang optimal. Selain itu, analisis probabilistik menunjukkan bahwa kualitas tidur (*Sleep Quality*) dan tingkat stres belajar (*Study Stress Level*) merupakan fitur yang paling berpengaruh dalam memprediksi risiko kesehatan mental. Temuan ini menunjukkan bahwa *Naive Bayes* efektif digunakan untuk klasifikasi *multi-class* dengan hasil yang dapat diinterpretasikan. Penelitian ini memberikan kontribusi terhadap upaya deteksi dini dan menjadi dasar untuk intervensi yang lebih tepat dalam pengelolaan kesehatan mental di lingkungan perguruan tinggi.

**Kata Kunci:** Kesehatan mental; Klasifikasi; Naive Bayes; Prediksi risiko; Analisis probabilistik

## Abstract

The mental health of university students is a growing concern as academic, emotional, and social pressures contribute to increased psychological risks. This study aims to classify mental health risk levels Low, Medium, and High among students using the Naive Bayes classification algorithm. A dataset consisting of 1,000 entries and 11 key variables was utilized, covering academic, psychological, and behavioral factors. The preprocessing stage included data cleaning, label encoding, normalization, and rule-based labeling to determine the target classes. Model training and testing were conducted using stratified data splitting to preserve class distribution. The initial model achieved a classification accuracy of 88,67%, with macro average F1-score of 0.87 and weighted average F1-score of 0.88. Grid Search optimization with k-fold cross-validation was applied but showed no significant improvement, indicating the model was already in optimal configuration. Furthermore, probabilistic analysis revealed that Sleep Quality and Study Stress Level were the most influential features in predicting mental health risks. The findings suggest that Naive Bayes is effective for multi-class classification with interpretable results. This research contributes to early detection efforts and offers a foundation for targeted interventions in university mental health management.

**Keywords:** Mental health; Classification; Naive Bayes; Risk prediction; Probabilistic analysis

## 1. PENDAHULUAN

Kesehatan Mental adalah keadaan pikiran seseorang yang menggambarkan karakter dirinya secara umum [1]. Kesehatan mental telah menjadi topik penting di dunia pendidikan khususnya di perguruan tinggi, karena memiliki peran yang signifikan dalam menunjang pencapaian akademik serta mendukung perkembangan mahasiswa secara keseluruhan [2]. Data Riset Kesehatan Dasar (Rikesdas) tahun 2018 menunjukkan bahwa lebih dari 19 juta penduduk yang berusia di atas 15 tahun mengalami depresi [3]. Ketika menelusuri prevalensi depresi di Indonesia, ditemukan bahwa tingkat depresi tertinggi dialami oleh individu dalam rentang usia remaja hingga dewasa muda dan cenderung berkurang seiring bertambahnya usia [4]. Berdasarkan hasil Indonesia *National Adolescent Mental Health Survei* (I-NAMHS) tahun 2022, terjadinya gangguan mental pada remaja Indonesia antara usia 10 hingga 17 tahun yang menunjukkan satu dari tiga remaja Indonesia memiliki masalah kesehatan mental. Hal ini setara dengan 15,5 juta remaja yang terdiagnosis sesuai dengan panduan *Diagnostic and Statistical Manual of Mental Disorders* Edisi Kelima (DSM-5) [5]. Kesehatan mental mahasiswa menjadi aspek penting yang perlu diperhatikan dalam lingkungan perguruan tinggi, masa perkuliahan merupakan periode peralihan yang signifikan dalam kehidupan individu, di mana pengalaman mahasiswa tidak hanya tentang prestasi akademik akan tetapi melibatkan berbagai tantangan mental, emosional dan sosial [6].

Seiring dengan berkembangnya teknologi, pendekatan berbasis *machine learning* semakin banyak digunakan untuk mendeteksi dan memprediksi kondisi kesehatan mental mahasiswa secara lebih dini dan akurat. Salah satu algoritma yang sering digunakan dalam klasifikasi data adalah *Naive Bayes* yang merupakan algoritma yang dapat mengklasifikasikan suatu variabel tertentu dengan metode probabilitas dan statistik [7]. Beberapa penelitian sebelumnya telah mengkaji prediksi kesehatan mental mahasiswa. Penelitian [5] memprediksi kesehatan mental mahasiswa menggunakan algoritma *Naive Bayes* dengan pendekatan SEMMA (*Sample, Explore, Modify, Model, and Assess*) pada dataset yang berisi 101 data, terdiri dari 9 variabel prediksi dan 1 variabel target, dengan hasil akurasi sebesar 93,33%. Selanjutnya, penelitian [6] mengklasifikasi kesehatan mental mahasiswa menggunakan *Naive Bayes* pada dataset yang berisi 100 data dengan hasil penelitian menunjukkan tingkat akurasi berbeda untuk setiap kategori prediksi, yaitu 70%

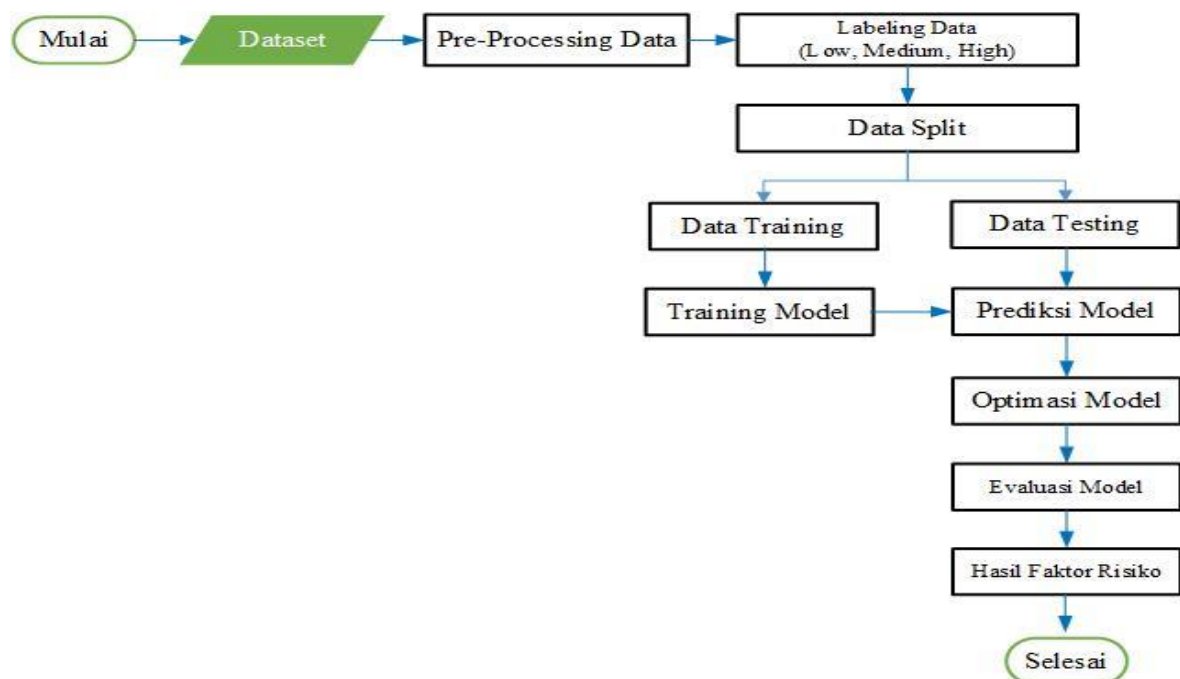
untuk depresi, 75% untuk kecemasan dan 75% untuk serangan panik. Sementara itu, penelitian [8] membandingkan performa algoritma *Naive Bayes Classifier* dan *Support Vector Machine* (SVM) dalam mengklasifikasi kesehatan mental mahasiswa menggunakan *dataset* yang memiliki 101 data dengan menunjukkan bahwa algoritma *Naive Bayes Classifier* menghasilkan akurasi 86,87%. Penelitian [9] mengembangkan model prediksi tingkat depresi pada mahasiswa dengan menggunakan algoritma *Naive Bayes* berbasis *machine learning*. Penelitian ini melibatkan 519 mahasiswa dan menghasilkan akurasi sebesar 78,03%. Meskipun demikian, penelitian ini masih terbatas pada jumlah variabel psikologis yang digunakan dan belum menguji performa model pada populasi luar. Selain itu, penelitian [10] menggunakan metode Tes Materialistik Aspek Sosial (TMAS) untuk klasifikasi kesehatan mental mahasiswa Universitas Global Jakarta, dengan membandingkan algoritma *Decision Tree*, *Random Forest*, dan *Logistic Regression*. Penelitian ini menggunakan jumlah sampel sebanyak 50 responden, dan menunjukkan bahwa *Logistic Regression* memberikan hasil terbaik dengan akurasi 90% dan *F1-score* sebesar 92%. Penelitian ini menyoroti potensi model klasifikasi sederhana dalam memprediksi kecemasan mahasiswa, namun masih memiliki keterbatasan dari sisi ukuran sampel dan jumlah variabel yang terbatas dari kuesioner TMAS.

Berdasarkan temuan dari penelitian sebelumnya, dapat disimpulkan bahwa sebagian besar penelitian masih menggunakan *dataset* berukuran terbatas dan hanya berfokus pada variabel psikologis, seperti depresi, kecemasan, dan serangan panik. Penggunaan variabel tambahan seperti stres akademik dan kualitas tidur masih jarang dilibatkan. Pendekatan pelabelan risiko umumnya belum dijelaskan secara eksplisit, serta belum banyak studi yang menyertakan analisis kontribusi masing-masing fitur terhadap hasil klasifikasi. Selain itu, performa metode *Naive Bayes* pada *dataset* yang lebih besar dan kompleks masih jarang dievaluasi. Tanpa pengujian pada data yang lebih beragam, model yang dihasilkan berisiko kurang representatif terhadap kondisi kesehatan mental mahasiswa secara menyeluruh. Hal ini menunjukkan bahwa masih terdapat celah penelitian (*research gap*) terkait pengembangan model klasifikasi risiko kesehatan mental mahasiswa yang lebih komprehensif dan akurat.

Mengacu pada celah tersebut, penelitian ini bertujuan untuk mengklasifikasikan risiko kesehatan mental mahasiswa ke dalam tiga kategori, yaitu rendah, sedang, dan tinggi, dengan menggunakan algoritma *Naive Bayes* pada *dataset* dengan jumlah data yang besar dan variabel yang beragam, serta pendekatan klasifikasi *multi-class*. Penelitian ini juga mengevaluasi performa model berdasarkan metrik akurasi, *precision*, *recall*, dan *F1-score*, serta menganalisis pengaruh relatif dari masing-masing fitur terhadap prediksi model, guna mengidentifikasi faktor-faktor risiko yang paling berpengaruh dalam memprediksi kesehatan mental mahasiswa melalui pendekatan berbasis probabilitistik dari model *Naive Bayes*. Untuk mendukung hasil klasifikasi yang optimal, dilakukan juga optimasi model dengan *tuning hyperparameter*. Diharapkan hasil penelitian ini dapat mengisi *gap* yang ada dan berkontribusi pada pengembangan sistem deteksi dini yang efisien, akurat, dan mudah diinterpretasikan di lingkungan perguruan tinggi.

## 2. METODOLOGI PENELITIAN

Penelitian ini mengikuti pendekatan berbasis *supervised learning*, di mana model dikembangkan melalui serangkaian tahapan sistematis. Diagram alur dalam penelitian ini menggambarkan proses yang mencakup seluruh tahapan dari pemrosesan awal hingga analisis hasil prediksi, seperti yang ditunjukkan pada gambar 1.



**Gambar 1.** Alur Penelitian

## 2.1 Dataset

Dataset yang digunakan dalam penelitian ini berasal dari situs *Kaggle*, dengan total 1.000 data dan 15 variabel, yang terdiri dari *Gender, Age, Course, Year of Study, CGPA, Depression, Anxiety, Panic Attack, Symptom Frequency Last 7 Days, Specialist Treatment, Has Mental Health Support, Sleep Quality, Study Stress Level, Study Hours Per Week, dan Academic Engagement*.

## 2.2 Preprocessing Dataset

Pada tahap pra-pemrosesan (*preprocessing*), dilakukan sejumlah langkah untuk menyiapkan data sebelum digunakan dalam proses pelatihan model klasifikasi. Tahap awal dimulai dengan pembersihan data (*data cleaning*) untuk menghapus data yang tidak lengkap atau mengandung nilai tidak valid, guna memastikan integritas dan kualitas data yang digunakan. Setelah itu, dilakukan transformasi terhadap variabel kategorikal menjadi numerik menggunakan metode *Label Encoding*. *Label Encoding* merupakan teknik konversi data kategorikal ke dalam format numerik dengan memberikan nilai angka unik untuk setiap kategori [11]. Dalam penelitian ini, digunakan teknik normalisasi untuk membuat skala nilai pada fitur-fitur numerik ke dalam rentang (0,1) menggunakan pendekatan *Min-Max Scaling* [12]. Langkah ini bertujuan untuk mencegah dominasi fitur dengan skala besar terhadap hasil pelatihan model, sehingga model dapat bekerja secara adil dan optimal dalam memproses seluruh fitur yang digunakan.

## 2.3 Pelabelan Dataset

Dalam penelitian ini, proses pelabelan data dilakukan menggunakan pendekatan *rule-based labeling*, yaitu metode penentuan label kelas berdasarkan aturan tertentu yang menggabungkan nilai dari beberapa variabel yang merepresentasikan kondisi kesehatan mental mahasiswa. Pendekatan ini efektif dalam mengidentifikasi pola risiko pada pasien depresi berdasarkan data historis dan skor psikologis, serta memberikan hasil yang mudah diinterpretasikan dalam konteks medis [13]. Selain itu, pendekatan *rule-based* tetap relevan dalam klasifikasi catatan kesehatan mental karena keunggulannya dalam hal transparansi dan kemudahan dalam dipahami [14]. Sementara itu, [15] menerapkan algoritma Apriori sebagai bagian dari pendekatan berbasis aturan (*rule-based*) untuk mengidentifikasi pola hubungan antar variabel risiko kesehatan mental mahasiswa, sebelum digunakan dalam model prediksi.

## 2.4 Pembagian Dataset

Pembagian *dataset* menggunakan teknik stratifikasi (*stratified split*) berbasis label dan distribusi fitur, yang dapat meningkatkan akurasi model dalam pengenalan aktivitas manusia dengan menjaga keseragaman distribusi antar sub set data [16]. Selain itu, studi dari [17] menunjukkan bahwa *stratified split* sampling efektif dalam menjaga representasi yang seimbang antar sub kelompok dalam data rekam medis elektronik, sehingga mendukung generalisasi model yang lebih baik.

## 2.5 Pelatihan Model

Pada tahap pelatihan (*training*), model *Naïve Bayes* dilatih menggunakan data latih untuk mempelajari distribusi probabilitas dari setiap fitur terhadap masing-masing kelas. Model mengasumsikan bahwa setiap fitur bekerja secara terpisah satu sama lain jika diketahui kelasnya, sehingga proses pelatihan menjadi lebih sederhana dan cepat. *Naive Bayes* merupakan algoritma klasifikasi *data mining* probabilistik sederhana yang menggunakan teknik probabilitas dengan menghitung frekuensi dan kombinasi nilai dalam suatu kumpulan data tertentu menggunakan teknik statistik [7], [18], [19]. Algoritma *Naive Bayes* adalah metode klasifikasi yang menggunakan probabilitas untuk menentukan klasifikasi dengan memprediksi peluang berdasarkan data sebelumnya, didasarkan pada *Teorema Bayes* yang menghitung probabilitas suatu hipotesis kasus berdasarkan data yang ada [20], [21].

Tujuan dari pelatihan ini adalah untuk membangun model probabilistik yang mampu mengenali pola hubungan antara fitur-fitur (misalnya, depresi, kecemasan, stres akademik) dengan kategori risiko kesehatan mental mahasiswa (Rendah, Sedang, Tinggi) dengan menggunakan metode klasifikasi *Naive Bayes* yang merupakan kumpulan algoritma klasifikasi yang menggunakan asumsi independensi antara fitur-fitur yang ada pada dataset, sehingga memungkinkan model untuk melakukan prediksi dengan efisien dan efektif [22]. *Naive Bayes* merupakan pendekatan yang sederhana namun memberikan hasil yang optimal dalam proses klasifikasi dan pengelompokan data, penggunaan algoritma ini sangat efisien dan memiliki tingkat akurasi yang tinggi [23]. Model *Naive Bayes* menghitung probabilitas suatu kelas  $H$  terhadap data fitur  $X$  menggunakan *Teorema Bayes* pada persamaan (1) [18].

$$P(H|X) = \frac{P(X|H) \times P(H)}{P(X)} \quad (1)$$

dengan,

- $X$  : data yang memiliki *class* dan belum diketahui,
- $H$  : hipotesis data  $X$ , merupakan *class* yang spesifik,
- $P(H)$  : probabilitas hipotesis  $H$  (*posteriori probability*),
- $P(H|X)$  : probabilitas hipotesis  $H$  mengikuti kondisi  $X$  (*posteriori probability*),
- $P(X|H)$  : probabilitas  $X$  mengikuti kondisi hipotesis  $H$ ,

$P(X)$  : probabilitas  $X$ .

## 2.6 Pengujian dan Evaluasi Model

Setelah model *Naïve Bayes* dilatih menggunakan data latih, tahap selanjutnya adalah pengujian model (*prediction model*) dengan menggunakan data uji. Pada tahap ini, model digunakan untuk memprediksi label risiko kesehatan mental berdasarkan data uji yang belum pernah dilihat sebelumnya. Hasil prediksi tersebut kemudian dibandingkan dengan label sebenarnya untuk menghitung metrik evaluasi. Evaluasi dilakukan menggunakan metrik evaluasi dalam klasifikasi, yaitu akurasi, *precision*, *recall*, dan *F1-score* yang masing-masing memiliki fungsi dalam mengukur efektivitas model [22]. Akurasi adalah ukuran kinerja model klasifikasi yang menunjukkan proporsi prediksi yang benar dari keseluruhan data [12], [24]. Akurasi dapat dihitung dengan persamaan (2).

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (2)$$

*Precision (Positive Predictive Value)* mengukur tingkat keakuratan model dalam memprediksi kelas positif [12], [24]. Presisi dapat dihitung dengan persamaan (3).

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

*Recall (True Positive Rate)* mengukur kemampuan model dalam menemukan semua data positif yang benar [12], [24]. *Recall* dapat dihitung dengan persamaan (4).

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

Sementara itu, *F1-score* merupakan rata-rata harmonik dari *precision* dan *recall*, yang digunakan untuk menyeimbangkan kedua metrik tersebut [12]. *F1-score* dapat dihitung dengan persamaan (5).

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (5)$$

Selain metrik evaluasi tersebut, digunakan pula *Confusion Matrix* untuk menggambarkan distribusi hasil prediksi model: *True Positive (TP)*, *False Positive (FP)*, *True Negative (TN)*, dan *False Negative (FN)* [22]. *TP* mengindikasikan jumlah data yang diklasifikasikan benar sebagai positif, *TN* adalah jumlah data yang benar diklasifikasikan sebagai negatif, *FP* menunjukkan data negatif yang salah diklasifikasikan sebagai positif, dan *FN* adalah data positif yang salah diklasifikasikan sebagai negatif [12].

## 2.7 Optimasi Model

Dalam penelitian ini, dilakukan proses optimasi model menggunakan teknik *Grid Search* dengan tujuan untuk mengevaluasi apakah hasil prediksi awal dari model *Naïve Bayes* masih dapat ditingkatkan. *Grid Search* merupakan metode sistematis yang menguji berbagai kombinasi nilai *hyperparameter* secara menyeluruh, guna menemukan konfigurasi terbaik yang menghasilkan kinerja model paling optimal [12].

Teknik ini umumnya dikombinasikan dengan *k-fold cross-validation*, yang berfungsi untuk mengevaluasi stabilitas performa model dengan membagi data latih menjadi beberapa sub set dan melakukan pelatihan serta pengujian secara berulang [25], [26]. Pendekatan ini menghasilkan estimasi performa model yang lebih andal karena mengurangi bias akibat pembagian data yang tidak merata. Setelah dilakukan evaluasi terhadap setiap kombinasi parameter, nilai akurasi rata-rata dari setiap konfigurasi dibandingkan untuk menentukan pengaturan *hyperparameter* terbaik [12].

## 2.8 Analisis Kontribusi Fitur

Setelah model dilatih dan dievaluasi, dilakukan analisis probabilistik terhadap fitur-fitur yang digunakan untuk mengetahui sejauh mana masing-masing fitur berkontribusi dalam menentukan klasifikasi risiko kesehatan mental. Analisis ini didasarkan pada parameter internal model *Naïve Bayes*, yaitu nilai rata-rata (*mean*) dan varians (*variance*) dari setiap fitur terhadap setiap kelas. Nilai *mean* digunakan untuk mengukur seberapa jauh perbedaan distribusi nilai suatu fitur antar kelas, sedangkan varians digunakan untuk menilai konsistensi nilai fitur dalam satu kelas. Rasio antara perbedaan *mean* dan rata-rata varians kemudian dihitung sebagai dasar untuk mengukur pengaruh relatif masing-masing fitur.

# 3. HASIL DAN PEMBAHASAN



Bagian ini menyajikan hasil dari penerapan metode klasifikasi terhadap data kesehatan mental mahasiswa. Proses yang dilakukan mencakup pemaparan karakteristik *dataset* awal, tahapan *preprocessing*, pembagian data latih dan uji, serta hasil klasifikasi menggunakan algoritma *Naïve Bayes*. Selain itu, disertakan pula evaluasi performa model, hasil optimasi menggunakan teknik *Grid Search*, serta analisis probabilistik terhadap faktor-faktor risiko yang berpengaruh.

### 3.1 Dataset

Dari total 15 variabel yang ada pada *dataset*, hanya 11 yang digunakan dalam proses klasifikasi. Variabel yang digunakan dapat dilihat pada tabel 1, yang masih merepresentasikan bentuk data mentah sebelum dilakukan proses *preprocessing*.

**Tabel 1.** Daftar Variabel yang Digunakan

No.	Variabel	Tipe Data	Keterangan
1.	<i>Gender</i>	<i>Object</i>	Jenis kelamin <i>Female</i> (Perempuan) dan <i>Male</i> (Laki-Laki)
2.	<i>Age</i>	<i>Integer</i>	Usia mahasiswa dalam rentang 18 – 25 tahun
3.	<i>Course</i>	<i>Object</i>	Nama program studi, terdiri dari 49 nama jurusan
4.	<i>Year of Study</i>	<i>Object</i>	Tahun studi, Year 1 – Year 4
5.	<i>CGPA</i>	<i>Float</i>	IPK dengan nilai antara 2.00 – 4.00
6.	<i>Depression</i>	<i>Integer</i>	0 = Tidak Mengalami 1 = Mengalami
7.	<i>Anxiety</i>	<i>Integer</i>	0 = Tidak Mengalami 1 = Mengalami
8.	<i>Panic Attack</i>	<i>Integer</i>	0 = Tidak Mengalami 1 = Mengalami
9.	<i>Sleep Quality</i>	<i>Integer</i>	1 = Sangat Buruk 2 = Buruk 3 = Netral 4 = Baik 5 = Sangat Baik
10.	<i>Study Stress Level</i>	<i>Integer</i>	1 = Tidak Stres 2 = Ringan 3 = Sedang 4 = Stres Berat 5 = Sangat Stres
11.	<i>Study Hours Per Week</i>	<i>Integer</i>	Lama waktu belajar per minggu, rentang 1 – 19 jam

Pemilihan variabel pada tabel 1 didasarkan pada fokus penelitian terhadap faktor-faktor risiko utama yang dapat memengaruhi kesehatan mental mahasiswa, baik dari aspek demografis, akademik, psikologis, maupun perilaku. Sementara itu, empat variabel lainnya yaitu *Symptom Frequency Last 7 Days*, *Specialist Treatment*, *Has Mental Health Support*, dan *Academic Engagement* tidak disertakan dalam proses klasifikasi, karena dianggap tidak merepresentasikan faktor risiko secara langsung. *Symptom Frequency Last 7 Days*, *Specialist Treatment*, dan *Has Mental Health Support* cenderung mencerminkan kondisi pasca-diagnosis atau bentuk intervensi, sementara *Academic Engagement* tidak digunakan karena tidak tersedia penjelasan yang memadai mengenai makna skala atau instrumen pengukuran tersebut.

### 3.2 Preprocessing

Pada tahap *preprocessing* dilakukan beberapa langkah, yaitu pembersihan data, transformasi data kategorikal ke numerik, dan normalisasi data. Hasil dari pembersihan data dan transformasi, dapat dilihat pada gambar 2.

	Gender	Age	YearOfStudy	CGPA	Depression	Anxiety	PanicAttack	SleepQuality	StudyStressLevel	StudyHoursPerWeek	course
0	0	24	3	2.38	1	0	0	4	5	8	7
1	0	18	3	4.00	0	1	0	4	4	13	7
2	0	25	3	3.68	0	0	1	1	2	13	7
3	0	18	4	4.00	0	0	0	5	1	19	9
4	0	20	4	2.00	1	1	0	2	4	3	9
...	...	...	...	...	...	...	...	...	...	...	...
995	0	20	1	2.31	0	1	0	1	3	7	9
996	0	23	1	3.72	1	0	0	4	1	12	9
997	0	19	2	2.00	0	0	1	5	4	3	3
998	0	24	2	3.89	1	0	1	5	4	13	3
999	0	20	2	2.46	1	0	0	2	5	8	3

**Gambar 2.** Hasil Pembersihan Data dan Transformasi

Berdasarkan gambar 2, pada tahap pembersihan data menemukan bahwa adanya penulisan nama program studi (*Course*) yang sama tetapi berbeda dalam huruf maupun kata. Pada tahap ini dilakukan penyesuaian penulisan nama program studi agar seragam dan dapat diproses pada tahap transformasi selanjutnya. Proses transformasi diterapkan pada beberapa variabel, yaitu *Gender*, *Year of Study*, serta *Course* yang semula berupa nama program studi diubah menjadi

nilai numerik berdasarkan pengelompokannya. Untuk variabel *Course*, dilakukan pengelompokan ke dalam 10 kelompok bidang keilmuan. Setiap kelompok *Course* kemudian diberi label angka berbeda untuk menyederhanakan struktur data dan mengurangi kompleksitas representasi. Setelah proses transformasi, dilakukan normalisasi pada variabel numerik yang memiliki skala nilai berbeda, seperti *Age*, *CGPA*, *StudyHoursPerWeek*, *SleepQuality*, dan *StudyStressLevel* guna menyetarakan kontribusi antar fitur dalam model. Tabel 2 menyajikan penjelasan singkat mengenai isi data dari hasil pembersihan data dan transformasi.

**Tabel 2.** Penjelasan Hasil Pembersihan dan Transformasi

No.	Variabel	Tipe Data	Keterangan
1.	<i>Gender</i>	<i>Integer</i>	0 = <i>Female</i> (Perempuan) 1 = <i>Male</i> (Laki-Laki)
2.	<i>Course</i>	<i>Integer</i>	Nama program studi, terdiri dari 38 nama jurusan yang dikelompokkan menjadi : 0 = Ekonomi dan Bisnis 1 = Hukum 2 = Humaniora 3 = Ilmu Sosial 4 = Kesehatan 5 = Lainnya 6 = Psikologi 7 = Sains dan Teknologi 8 = Studi Islam 9 = Teknik
3.	<i>Year of Study</i>	<i>Integer</i>	Tahun studi, 1 – 4

### 3.3 Pelabelan Dataset

Penentuan label risiko dilakukan dengan menghitung skor total dari lima variabel utama yang memengaruhi kondisi mental, yaitu *Depression*, *Anxiety*, *Panic Attack*, *Study Stress Level*, dan *Reversed Sleep Quality*. Variabel *Sleep Quality* diubah menjadi *Reversed Sleep Quality* untuk menyelaraskan arah interpretasi risiko semakin tinggi nilai, semakin besar tingkat risikonya. Skor total kemudian digunakan untuk mengelompokkan data ke dalam tiga kategori: Rendah, Sedang, dan Tinggi, sebagaimana ditunjukkan pada gambar 3.

TotalRiskScore	
count	1000.000000
mean	7.477000
std	2.227652
min	2.000000
25%	6.000000
50%	7.000000
75%	9.000000
max	13.000000

**Gambar 3.** Skor Total Risiko

Skor total risiko sebagaimana ditampilkan pada gambar 3 memiliki nilai minimum sebesar 2 dan maksimum 13, dengan nilai kuartil pertama (Q1) sebesar 6 dan kuartil ketiga (Q3) sebesar 9. Berdasarkan distribusi ini, dilakukan klasifikasi skor ke dalam tiga kategori: Rendah ( $\leq 6$ ), Sedang ( $> 6$  hingga  $\leq 9$ ), dan Tinggi ( $> 9$ ). Hasil pelabelan data dapat dilihat pada gambar 4.

Depression	Anxiety	Panicattack	StudyStressLevel	SleepQuality	ReversedSleepQuality	TotalRiskScore	RiskLabel	RiskLabelNum
1	0	0	5	4	2	8	Sedang	1
0	1	0	4	4	2	7	Sedang	1
0	0	1	2	1	5	8	Sedang	1
0	0	0	1	5	1	2	Rendah	0
1	1	0	4	2	4	10	Tinggi	2
0	0	1	1	4	2	4	Rendah	0

**Gambar 4.** Hasil Pelabelan Data

Hasil pelabelan data pada gambar 4 menunjukkan distribusi label ke dalam tiga kategori: Rendah, Sedang, dan Tinggi, dengan hasil distribusi label yang kurang seimbang terutama pada kategori Tinggi yang memiliki jumlah distribusi paling sedikit. Rincian distribusi jumlah data berdasarkan label risiko kesehatan mental yang telah ditetapkan dapat dilihat pada tabel 3. Label ini diperoleh melalui proses pelabelan berbasis aturan (*rule-based labeling*) yang membagi skor total risiko ke dalam tiga kategori utama.

**Tabel 3.** Distribusi Risiko Label

Risiko Label Num	Risiko Label	Distribusi Label
0	Rendah	468
1	Sedang	342
2	Tinggi	190

Berdasarkan tabel 3, mayoritas data termasuk dalam kategori risiko sedang, diikuti oleh kategori rendah, dan paling sedikit pada kategori tinggi. Hal ini menunjukkan bahwa sebagian besar mahasiswa dalam *dataset* berada pada kondisi mental yang belum optimal, namun belum masuk kategori risiko tinggi. Ketidakseimbangan jumlah data antar kategori ini juga menjadi pertimbangan dalam evaluasi performa model klasifikasi.

### 3.4 Pembagian Dataset

Pembagian *dataset* dalam penelitian ini dilakukan menjadi dua bagian, yaitu data latih sebesar 70% dan data uji sebesar 30% dari total 1.000 data. Untuk menjaga distribusi label risiko kesehatan mental tetap proporsional pada kedua sub set data, digunakan teknik *Stratified Split*. Teknik ini memastikan bahwa setiap kelas (Rendah, Sedang, dan Tinggi) terwakili secara seimbang, sehingga evaluasi model menjadi lebih adil, tidak bias terhadap kelas mayoritas, serta representatif terhadap distribusi populasi. Distribusi hasil pembagian data berdasarkan label dapat dilihat pada tabel 4.

**Tabel 4.** Hasil Pembagian Data

Label	Jumlah Data Latih	Jumlah Data Uji
Rendah (0)	239	103
Sedang (1)	328	140
Tinggi (2)	133	57

Distribusi pada tabel 4 menunjukkan bahwa teknik pembagian data telah berhasil mempertahankan proporsi label yang konsisten antara data latih dan data uji. Kelas Sedang merupakan mayoritas dalam *dataset*, sedangkan kelas Tinggi termasuk minoritas. Keberadaan ketidakseimbangan jumlah antar kelas ini menjadi faktor penting yang perlu diperhatikan dalam evaluasi model, khususnya pada metrik *recall* dan *F1-score* untuk kelas minoritas.

### 3.5 Klasifikasi Naive Bayes

Proses klasifikasi dilakukan dengan menerapkan algoritma *Naive Bayes* terhadap data uji yang telah dibagi sebelumnya menggunakan metode *Stratified Split*. Model dilatih pada data latih dan dilakukan prediksi pada data uji. Hasil evaluasi klasifikasi dapat dilihat pada tabel 5.

**Tabel 5.** Hasil Klasifikasi Awal

	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>	<i>Support</i>
Rendah	1.00	0.84	0.92	103
Sedang	0.80	1.00	0.89	140
Tinggi	1.00	0.68	0.81	57
<i>Accuracy</i>			0.89	300
<i>Macro Avg</i>	0.93	0.84	0.87	300
<i>Weighted Avg</i>	0.91	0.89	0.88	300

Hasil klasifikasi awal pada tabel 5 menunjukkan bahwa model *Naive Bayes* mampu mengklasifikasikan risiko kesehatan mental mahasiswa ke dalam tiga kategori dengan akurasi keseluruhan sebesar 89%. Nilai *precision*, *recall*, dan *f1-score* pada masing-masing kelas menunjukkan bahwa model bekerja cukup efektif dalam mengenali pola-pola risiko yang ada. Model menunjukkan performa paling tinggi pada kelas Sedang dengan *recall* sebesar 1.00, menandakan bahwa seluruh data kategori Sedang berhasil dikenali secara benar. Namun, *precision* berada pada angka 0.80, yang berarti sebagian prediksi kelas Sedang berasal dari kelas lain. Hal ini menunjukkan kecenderungan model untuk mengklasifikasikan data ke dalam kelas Sedang secara berlebihan.

Untuk kelas Rendah, model memperoleh *precision* sempurna (1.00) dan *f1-score* tertinggi (0.92), dengan *recall* 0.84. Artinya, semua prediksi kelas Rendah benar, tetapi tidak semua data kelas Rendah berhasil dikenali. Sementara itu, pada kelas Tinggi, meskipun *precision* juga mencapai 1.00, nilai *recall* hanya 0.68. Hal ini menunjukkan bahwa model cukup konservatif dalam mengklasifikasikan data sebagai risiko tinggi, yang berdampak pada terlewatnya sebagian data berisiko tinggi (*false negative*). Nilai *macro average* dan *weighted average* dari *f1-score*, masing-masing 0.87 dan 0.88, menunjukkan bahwa model memberikan kinerja seimbang antar kelas, meskipun masih perlu peningkatan pada kelas minoritas (Tinggi), yang memiliki jumlah data lebih sedikit. Secara keseluruhan, hasil ini menunjukkan bahwa metode *Naive Bayes* efektif digunakan untuk klasifikasi *multi-class* dalam konteks risiko kesehatan mental mahasiswa. Secara keseluruhan, *Naive Bayes* terbukti efektif untuk klasifikasi *multi-class* dalam konteks ini, dengan kompleksitas yang rendah.

### 3.6 Optimasi Model *Grid Search*

Proses optimasi model dilakukan menggunakan teknik *Grid Search* untuk mengevaluasi apakah hasil prediksi awal dari model Naïve Bayes masih dapat ditingkatkan melalui penyesuaian parameter. Optimasi ini dilakukan secara independen dari performa awal, dengan dua tujuan utama. Jika hasil awal sudah optimal, maka tuning berfungsi sebagai validasi tambahan untuk memastikan efisiensi model. Sebaliknya, jika hasil awal belum maksimal, proses tuning dapat membantu meningkatkan akurasi model melalui konfigurasi parameter terbaik. Oleh karena itu, optimasi model tidak hanya sebagai pelengkap, tetapi juga merupakan strategi evaluatif penting dalam pengembangan model klasifikasi yang akurat. Hasil perbandingan performa model sebelum dan sesudah *tuning* dapat dilihat pada gambar 5.

Default Model:				
	precision	recall	f1-score	support
Rendah	1.00	0.84	0.92	103
Sedang	0.80	1.00	0.89	140
Tinggi	1.00	0.68	0.81	57
accuracy			0.89	300
macro avg	0.93	0.84	0.87	300
weighted avg	0.91	0.89	0.88	300
Tuned Model:				
	precision	recall	f1-score	support
Rendah	1.00	0.84	0.92	103
Sedang	0.80	1.00	0.89	140
Tinggi	1.00	0.68	0.81	57
accuracy			0.89	300
macro avg	0.93	0.84	0.87	300
weighted avg	0.91	0.89	0.88	300

**Gambar 5.** Perbandingan Hasil Awal dan Hasil Optimasi

Gambar 5 menunjukkan bahwa nilai metrik evaluasi pada model *default* (awal) dan model hasil *tuning* adalah sama persis, dengan akurasi keseluruhan sebesar 0.89, *macro average F1-score* sebesar 0.87, dan *weighted average F1-score* sebesar 0.88. Nilai *precision* dan *recall* pada masing-masing kelas juga tidak mengalami perubahan. Temuan ini menunjukkan bahwa optimasi parameter menggunakan *Grid Search* tidak memberikan peningkatan performa terhadap model Naïve Bayes yang digunakan dalam penelitian ini. Hal ini menandakan bahwa model awal sudah berada dalam konfigurasi parameter yang optimal, atau bahwa parameter Naïve Bayes memang tidak terlalu sensitif terhadap *tuning* pada *dataset* ini. Performa awal model dapat dianggap cukup efisien dan stabil untuk digunakan dalam konteks klasifikasi risiko kesehatan mental mahasiswa.

### 3.7 Evaluasi Model

Evaluasi performa model dilakukan dengan menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score*. Metrik ini digunakan untuk mengukur seberapa baik model mengenali masing-masing kategori risiko secara proporsional dan menyeluruh. Rangkuman hasil evaluasi model ditampilkan pada tabel 6.

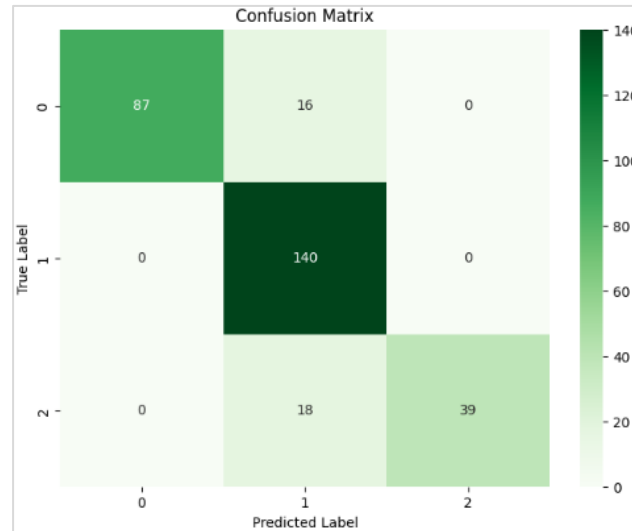
**Tabel 6.** Metrik Evaluasi

Metrik	Nilai
<i>Accuracy</i>	88.67%
<i>Precision</i>	90.88%
<i>Recall</i>	88.67%
<i>F1-score</i>	88.49%

Berdasarkan tabel 6, model mencapai akurasi keseluruhan sebesar 88,67%, dengan *precision* 90,88% dan *f1-score* 88,49%. Nilai *precision* yang cukup tinggi menunjukkan bahwa sebagian besar prediksi yang dilakukan oleh model adalah benar, sedangkan nilai *recall* yang seimbang mencerminkan kemampuan model dalam mendeteksi seluruh data relevan dari setiap kategori risiko.

Sebagai pelengkap evaluasi, digunakan *confusion matrix* untuk menggambarkan distribusi prediksi model terhadap label aktual. Visualisasi *confusion matrix* disajikan pada gambar 6.





Gambar 6. Confusion Matrix

Berdasarkan gambar 6, diketahui bahwa kelas Sedang (label 1) diklasifikasikan dengan sangat baik, dengan 140 data dikenali secara benar tanpa kesalahan (*true positive*). Pada kelas Rendah (label 0), sebanyak 87 data berhasil diklasifikasikan dengan benar, sementara 16 data salah diklasifikasikan sebagai kelas Sedang. Hal ini menunjukkan bahwa model sedikit kesulitan membedakan karakteristik antara risiko rendah dan sedang.

Sementara itu, untuk kelas Tinggi (label 2), sebanyak 39 data berhasil dikenali secara benar, tetapi 18 data diklasifikasikan sebagai kelas Sedang. Hal ini mengindikasikan bahwa model memiliki kecenderungan untuk memetakan data ke kelas mayoritas (Sedang), terutama saat harus mengenali kelas minoritas seperti Tinggi. Hasil ini menegaskan pentingnya penanganan *class imbalance* dalam pengembangan model klasifikasi *multi-class*.

### 3.8 Analisis Kontribusi Fitur

Analisis ini bertujuan mengidentifikasi fitur yang paling berpengaruh terhadap klasifikasi risiko kesehatan mental mahasiswa. Metode yang digunakan berdasarkan distribusi probabilistik model *Naïve Bayes*, dengan menghitung rasio selisih rata-rata antar kelas terhadap rata-rata varians tiap fitur. Semakin tinggi nilai rasio, semakin besar pengaruh fitur dalam membedakan kelas risiko. Hasil analisis probabilistik terhadap masing-masing fitur ditunjukkan pada gambar 7, yang menampilkan urutan fitur berdasarkan tingkat pengaruhnya terhadap proses klasifikasi risiko kesehatan mental.

No.	Fitur	Skor Pengaruh
0	9 SleepQuality	2.098802
1	10 StudyStressLevel	1.966416
2	6 Depression	1.518913
3	8 PanicAttack	1.339418
4	7 Anxiety	1.336407
5	1 Gender	0.100201
6	4 YearOfStudy	0.073234
7	2 Age	0.044609
8	5 CGPA	0.026033
9	3 Course	0.021194
10	11 StudyHoursPerWeek	0.006773

Gambar 7. Hasil Faktor Risiko Pengaruh

Berdasarkan gambar 7, dapat disimpulkan bahwa fitur *Sleep Quality* dan *Study Stress Level* merupakan dua faktor risiko yang paling dominan, dengan skor pengaruh masing-masing sebesar 2.09 dan 1.96. Hal ini menunjukkan bahwa mahasiswa dengan kualitas tidur buruk atau tingkat stres akademik tinggi cenderung lebih rentan mengalami risiko kesehatan mental. Diikuti oleh *Depression*, *Panic Attack*, dan *Anxiety*, yang juga berkontribusi kuat terhadap klasifikasi.

Sementara itu, fitur-fitur seperti CGPA (IPK), *Course*, dan *Study Hours Per Week* memiliki skor pengaruh yang jauh lebih rendah, menandakan bahwa dalam konteks *dataset* ini, faktor akademik kuantitatif seperti nilai IPK dan durasi belajar tidak terlalu membedakan kondisi risiko mental secara signifikan. Temuan ini memperkuat relevansi pendekatan psikologis dan perilaku dalam prediksi kesehatan mental berbasis data.

## 4. KESIMPULAN

Penelitian ini berhasil mengimplementasikan algoritma *Naïve Bayes* untuk mengklasifikasikan risiko kesehatan mental mahasiswa ke dalam tiga kategori, yaitu Rendah, Sedang, dan Tinggi. Proses klasifikasi dilakukan menggunakan *dataset* berjumlah 1.000 data dengan 11 variabel utama yang mencerminkan faktor demografis, akademik, psikologis, dan perilaku mahasiswa. Model *Naïve Bayes* menunjukkan performa klasifikasi yang cukup baik, dengan akurasi sebesar 88,67%, nilai *macro average F1-score* sebesar 0.87, dan *weighted average F1-score* sebesar 0.88. Hasil tersebut menunjukkan bahwa model bekerja secara seimbang dalam mengenali ketiga kategori risiko. Proses optimasi menggunakan *Grid Search* dan *k-fold cross-validation* menunjukkan bahwa parameter awal model telah berada pada konfigurasi yang optimal, karena tidak terjadi peningkatan performa setelah *tuning* dilakukan. Selain itu, melalui pendekatan analisis probabilistik terhadap distribusi parameter internal model, ditemukan bahwa *Sleep Quality* dan *Study Stress Level* merupakan fitur yang paling dominan dalam memengaruhi hasil klasifikasi, diikuti oleh *Depression*, *Panic Attack*, dan *Anxiety*. Kelebihan dari penelitian ini adalah penggunaan pendekatan *rule-based* dalam pelabelan data yang transparan dan mudah dipahami, serta pendekatan probabilistik yang digunakan untuk mengevaluasi pengaruh fitur. Namun, kelemahan model terlihat pada performa kelas minoritas (Tinggi) yang masih rendah akibat ketidakseimbangan jumlah data antar kategori. Hal ini menunjukkan bahwa algoritma sederhana seperti *Naïve Bayes* tetap mampu memberikan hasil kompetitif jika didukung oleh strategi *labeling* dan analisis fitur yang tepat.

## REFERENCES

- [1] V. Lajjawala, A. Achaliya, H. Jatta, and V. Pinjarkar, "Classification Algorithms based Mental Health Prediction using Data Mining," no. Icces, pp. 1174–1178, 2020, doi: 10.1109/icces48766.2020.9137856.
- [2] F. Pratama, Z. Hadryan Nst, Z. Khairi, and L. Efrizoni, "Perbandingan Algoritma Random Forest Dan K-Nearest Neighbor Dalam Klasifikasi Kesehatan Mental Mahasiswa," *J. Ilm. Betrik*, vol. 15, no. 1, pp. 31–37, 2024.
- [3] T. Tanugeraha, A. J. Santoso, and S. P. Adithama, "Pengelompokan Gangguan Kesehatan Mental Mahasiswa yang sedang Menempuh Skripsi dengan Metode Fuzzy Mamdani," *J. Inform. Atma Jogja*, vol. 4, no. 1, pp. 77–84, 2023, doi: 10.24002/jiaj.v4i1.7445.
- [4] K. Peltzer and S. Pengpid, "High prevalence of depressive symptoms in a national sample of adults in Indonesia: Childhood adversity, sociodemographic factors and health risk behaviour," *Asian J. Psychiatr.*, vol. 33, no. March, pp. 52–59, 2018, doi: 10.1016/j.ajp.2018.03.017.
- [5] A. Salsabila Juwita, A. Rizky Kurniawan, A. Aryaputra Ashari, D. Tyan Putro, V. Nurcahyawati, and H. Artikel, "Implementasi Data Mining untuk Memprediksi Kesehatan Mental Mahasiswa menggunakan Algoritma Naïve Bayes," 2024, [Online]. Available: <http://studentjournal.umpo.ac.id/index.php/komputek>.
- [6] N. Moningka, M. Hafidurrohman, and W. R. Ajri Tri, "Klasifikasi Mental Mahasiswa Menggunakan Metode Machine Learning," *J. Quancom*, vol. 1, no. 2, pp. 27–32, 2023, [Online]. Available: <https://www.kaggle.com/datasets/shariful07/student-mental->
- [7] D. Sindy, "Analisis Sentimen Pada Komentar Mengenai Kartu Indonesia Pintar Menggunakan Metode Naïve Bayes," *J. Ris. Komput.*, vol. 12, no. 2, 2025, doi: 10.30865/jurikom.v12i2.8528.
- [8] H. D. Putra, L. Khairani, and D. Hastari, "Perbandingan Algoritma Naive Bayes Classifier dan Support Vector Machine untuk Klasifikasi Data Kesehatan Mental Mahasiswa," *SENTIMAS Semin. Nas. Penelit. dan Pengabd. Masy.*, pp. 120–125, 2023, [Online]. Available: <https://journal.irpi.or.id/index.php/sentimas>.
- [9] F. Torres Cruz, "Prediction of Depression Level in University Students through a Naive Bayes based Machine Learning Model," 2023, doi: <https://doi.org/10.48550/arXiv.2307.14371>.
- [10] A. A. Hapsari, A. S. Nursuwanda, H. Zuhriyah, and D. J. Vresdian, "Klasifikasi Kesehatan Mental Mahasiswa Model TMAS dengan Algoritma Decision Tree , Logistic Regression , dan Random Forest," *J. INTEK*, vol. 7, no. November, pp. 55–64, 2024, doi: <https://doi.org/10.37729/intek.v7i2.5690>.
- [11] C. Herdian, A. Kamila, and I. G. Agung Musa Budidarma, "Studi Kasus Feature Engineering Untuk Data Teks: Perbandingan Label Encoding dan One-Hot Encoding Pada Metode Linear Regresi," *Technol. J. Ilm.*, vol. 15, no. 1, p. 93, 2024, doi: 10.31602/tji.v15i1.13457.
- [12] A. Toha, P. Purwono, and W. Gata, "Model Prediksi Kualitas Udara dengan Support Vector Machines dengan Optimasi Hyperparameter GridSearch CV," *Bul. Ilm. Sarj. Tek. Elektro*, vol. 4, no. 1, pp. 12–21, 2022, doi: 10.12928/biste.v4i1.6079.
- [13] Y. Lin, S. Huang, G. E. Simon, and S. Liu, "Data-based Decision Rules to Personalize Depression Follow-up," *Sci. Rep.*, vol. 8, no. 1, pp. 4–11, 2018, doi: 10.1038/s41598-018-23326-1.
- [14] E. Rijcken, K. Zervanou, P. Mosteiro, F. Scheepers, M. Spruit, and U. Kaymak, "Machine Learning vs. Rule-Based Methods for Document Classification of Electronic Health Records within Mental Health Care - A Systematic Literature Review," *Nat. Lang. Process. J.*, vol. 10, no. January, pp. 0–15, 2024, doi: 10.1016/j.nlp.2025.100129.
- [15] Y. Fu, F. Ren, and J. Lin, "Apriori algorithm based prediction of students' mental health risks in the context of artificial intelligence," *Front. Public Heal.*, vol. 13, no. February, 2025, doi: 10.3389/fpubh.2025.1533934.
- [16] N. Trung Tuyen, N. Thanh Dang, D. Hon Minh, and T. Q. Nguyen, "Enhancing human activity recognition model performance through KMeans-based stratified data splitting," *Meas. Control (United Kingdom)*, 2025, doi: 10.1177/00202940241312873.
- [17] T. Huo, D. H. Glueck, E. A. Shenkman, and K. E. Muller, "Stratified split sampling of electronic health records," *BMC Med. Res. Methodol.*, vol. 23, no. 1, pp. 1–7, 2023, doi: 10.1186/s12874-023-01938-0.
- [18] L. Susanti, "Klasifikasi Tingkat Stress pada Mahasiswa Teknik Informatika dalam Melakukan Perkuliahan Metode Hybrid Menggunakan Algoritma Naive Bayes," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.)*, vol. 8, no. 3, p. 243, 2024, doi: 10.30998/string.v8i3.17096.
- [19] A. Rahmadani *et al.*, "Tes Gangguan Depresi , Kecemasan , Dan Stres Pada Mahasiswa Selama Masa Covid- 19 Menggunakan Algoritma Naïve Bayes Testing Depression , Anxiety , And Stress Disorders In Students During The Covid-19 Pandemic Using Naïve Bayes Algorithm," vol. 9, no. 3, pp. 1321–1330, 2022.

- [20] W. S. Wisnugraha, I. N. Farida, and ..., "Implementasi Algoritma Naïve Bayes Dalam Menentukan Diagnosa Tingkat Depresi Mahasiswa Akhir Terhadap Pengerjaan Skripsi," *Pros. SEMNAS ...*, vol. 7, pp. 919–928, 2023, [Online]. Available: <https://proceeding.unpkediri.ac.id/index.php/inotek/article/view/3517%0Ahttps://proceeding.unpkediri.ac.id/index.php/inotek/article/download/3517/2321>.
- [21] R. Alfarezy, E. Ermatita, and R. M. B. Wadu, "Implementasi Algoritma Naïve Bayes Untuk Analisis Klasifikasi Survei Kesehatan Mental (Studi Kasus: Open Sourcing Mental Illness)," *Inform. J. Ilmu Komput.*, vol. 19, no. 1, pp. 1–10, 2023, doi: 10.52958/iftk.v19i1.4696.
- [22] M. F. Eina, Y. H. Chrisnanto, and M. Melina, "Klasifikasi Telemarketing Menggunakan Naïve Bayes Classification Dan Wrapper Sequential Feature Selection," *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 7, no. 4, pp. 1189–1198, 2024, doi: 10.31539/intecom.v7i4.10846.
- [23] Q. A'yuniyah *et al.*, "Implementasi Algoritma Naïve Bayes Classifier (NBC) untuk Klasifikasi Penyakit Ginjal Kronik," *J. Sist. Komput. dan Inform.*, vol. 4, no. 1, p. 72, 2022, doi: 10.30865/json.v4i1.4781.
- [24] N. Mahfudza and M. Ikhsan, "Sentiment Analysis of Youtube Comments on Indonesian Presidential Candidates in 2024 using Naïve Bayes Classifier Method," *J. Rits. Komput.*, vol. 12, no. 2, pp. 140–148, 2025, doi: 10.30865/jurikom.v12i2.8538.
- [25] T. Yan, S. L. Shen, A. Zhou, and X. Chen, "Prediction of geological characteristics from shield operational parameters by integrating grid search and K-fold cross validation into stacking classification algorithm," *J. Rock Mech. Geotech. Eng.*, vol. 14, no. 4, pp. 1292–1303, 2022, doi: 10.1016/j.jrmge.2022.03.002.
- [26] K. R. Singh, K. P. Neethu, K. Madhurekaa, A. Harita, and P. Mohan, "Parallel SVM model for forest fire prediction," *Soft Comput. Lett.*, vol. 3, no. June, p. 100014, 2021, doi: 10.1016/j.socl.2021.100014.