



PERBANDINGAN ALGORITMA *NAIVE BAYES* DAN *MODIFIED NAIVE BAYES* DALAM KLASIFIKASI TINGKAT STRES MAHASISWA

Irwansyahputra¹, Nurdin²

^{1,2}Magister Teknologi Informasi, Universitas Malikussaleh

Email: irwansyahputra.257110201003@mhs.unimal.ac.id¹, nurdin@unimal.ac.id²

Informasi Artikel	ABSTRACT
<p>Riwayat artikel : Disubmit : 23 Mei 2026 Direvisi : 5 Juni 2026 Diterima : 13 Juni 2026 Dipublikasi : 30 Juni 2026</p> <p>Keywords: <i>student stress classification, machine learning, Naive Bayes, Laplace Corrected Naive Bayes, data mining</i></p>	<p><i>Student stress is a significant issue that can affect academic performance and mental health. Various studies have applied machine learning algorithms for stress classification; however, studies comparing Naive Bayes and Modified Naive Bayes (Laplace Corrected Naive Bayes) remain limited. This study aims to compare the performance of both algorithms in classifying student stress levels based on psychological and academic factors. The dataset used was Student Stress Factors: A Comprehensive Analysis from Kaggle, consisting of 1,100 records with 20 predictor attributes and 1 target attribute (stress level). The dataset was divided into training and testing sets using an 80:20 ratio. The research stages included data preprocessing, classification using Naive Bayes and Modified Naive Bayes, and evaluation using accuracy, precision, recall, and F1-score. The results showed that Naive Bayes achieved an accuracy of 84%, precision of 83%, recall of 82%, and F1-score of 82%, while Modified Naive Bayes achieved an accuracy of 89%, precision of 88%, recall of 87%, and F1-score of 87%. These findings indicate that Laplace Correction improves classification performance by reducing the zero-probability problem. Therefore, Modified Naive Bayes is more effective for student stress classification.</i></p>
<p>Kata Kunci: klasifikasi tingkat stres, machine learning, Naive Bayes, Laplace Corrected Naive Bayes, data mining</p>	<p style="text-align: center;">ABSTRAK</p> <p>Tingkat stres mahasiswa merupakan salah satu permasalahan yang dapat memengaruhi prestasi akademik dan kesehatan mental. Berbagai penelitian telah menerapkan algoritma machine learning untuk klasifikasi tingkat stres, namun penelitian yang membandingkan Naive Bayes dan Modified Naive Bayes (Laplace Corrected Naive Bayes) masih terbatas. Penelitian ini bertujuan membandingkan performa kedua algoritma dalam klasifikasi tingkat stres mahasiswa berdasarkan faktor psikologis dan akademik. Dataset yang digunakan adalah Student Stress Factors: A Comprehensive Analysis dari Kaggle yang terdiri atas 1.100 data dengan 20 atribut prediktor dan 1 atribut target (stress level). Data dibagi menjadi data training dan testing dengan rasio 80:20. Tahapan penelitian meliputi preprocessing data, klasifikasi menggunakan Naive Bayes dan Modified Naive Bayes, serta evaluasi menggunakan accuracy, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa Naive Bayes memperoleh accuracy 84%, precision 83%, recall 82%, dan F1-score 82%, sedangkan Modified Naive Bayes memperoleh accuracy 89%, precision 88%, recall 87%, dan F1-score 87%. Hasil tersebut menunjukkan bahwa penerapan Laplace Correction mampu meningkatkan performa klasifikasi dengan mengurangi permasalahan zero probability. Dengan demikian, Modified Naive Bayes lebih efektif digunakan untuk klasifikasi tingkat stres mahasiswa.</p>





PENDAHULUAN

Perkembangan teknologi informasi pada era digital telah menghasilkan data dalam jumlah yang sangat besar dari berbagai bidang. Kondisi tersebut mendorong pemanfaatan data mining sebagai suatu pendekatan untuk mengekstraksi pola, hubungan, dan pengetahuan yang tersembunyi dari sekumpulan data sehingga dapat digunakan sebagai dasar pengambilan keputusan yang lebih efektif. Data mining memanfaatkan berbagai teknik analisis data, baik yang berbasis statistik maupun machine learning, untuk menemukan informasi yang bernilai dari data yang tersedia (Maulita & Nurdin, 2023). Seiring perkembangannya, machine learning menjadi salah satu pendekatan yang banyak digunakan dalam data mining karena mampu mempelajari pola data secara otomatis dan menghasilkan model yang memiliki tingkat akurasi yang baik dalam proses prediksi maupun klasifikasi (Intan et al., 2021). Salah satu tugas utama dalam data mining adalah klasifikasi, yaitu proses mengelompokkan data ke dalam kelas atau kategori tertentu berdasarkan karakteristik yang dimiliki (Nurdin et al., 2021).

Mahasiswa merupakan salah satu kelompok yang rentan mengalami stres akibat berbagai tuntutan akademik maupun non-akademik, seperti beban tugas, ujian, persaingan akademik, manajemen waktu, serta penyesuaian terhadap lingkungan sosial. Tingkat stres yang tinggi dapat berdampak pada penurunan konsentrasi belajar, performa akademik, kualitas tidur, produktivitas, serta kesehatan mental mahasiswa secara keseluruhan (Bawuna et al., 2017). Oleh karena itu, diperlukan suatu pendekatan yang mampu mengidentifikasi tingkat stres mahasiswa secara cepat, objektif, dan akurat untuk mendukung upaya deteksi dini serta pengambilan keputusan yang tepat. Berbagai penelitian menunjukkan bahwa algoritma klasifikasi mampu memberikan performa yang baik dalam mengidentifikasi tingkat stres, seperti penggunaan algoritma Random Forest yang menghasilkan akurasi sebesar 93,65% pada klasifikasi tingkat stres berdasarkan parameter fisiologis saat tidur (Astari et al., 2023), serta penerapan metode klasifikasi berbasis machine learning yang efektif dalam meningkatkan akurasi prediksi dan peningkatan performa melalui modifikasi (Hidayat et al., 2020). Salah satu algoritma klasifikasi yang banyak digunakan adalah Naive Bayes karena memiliki struktur yang sederhana, proses komputasi yang cepat, serta mampu bekerja dengan baik pada data berukuran kecil maupun besar. Namun, asumsi independensi antar atribut yang digunakan oleh Naive Bayes sering kali tidak sepenuhnya sesuai dengan kondisi data nyata sehingga dapat memengaruhi kinerja klasifikasi. Oleh karena itu, berbagai pengembangan dilakukan melalui Modified Naive Bayes untuk meningkatkan kemampuan model dalam menangani hubungan antar atribut dan menghasilkan performa klasifikasi yang lebih baik dibandingkan Naive Bayes konvensional (Hidayat et al., 2020).

Penelitian mengenai klasifikasi tingkat stres mahasiswa telah banyak dilakukan dengan memanfaatkan teknik data mining dan machine learning. Susanti (2024) menerapkan algoritma Naive





Bayes untuk mengklasifikasikan tingkat stres mahasiswa Teknik Informatika dalam sistem pembelajaran hybrid. Hasil penelitian menunjukkan bahwa algoritma Naive Bayes mampu menghasilkan accuracy sebesar 80,00%, precision sebesar 77,78%, dan recall sebesar 77,78%, sehingga dinilai cukup efektif dalam mengidentifikasi tingkat stres mahasiswa (Susanti, 2024).

Penelitian lain dilakukan oleh Renaningtias et al. (2024) yang membandingkan algoritma Decision Tree C4.5 dan K-Nearest Neighbor (KNN) dalam klasifikasi masa studi dan tingkat stres mahasiswa. Hasil penelitian menunjukkan bahwa algoritma KNN memberikan performa yang lebih baik dibandingkan Decision Tree C4.5 dengan tingkat akurasi sebesar 64,17% pada klasifikasi tingkat stres mahasiswa (Renaningtias et al., 2024).

Anjani dan Yamasari (2023) melakukan klasifikasi tingkat stres mahasiswa menggunakan beberapa metode berbasis tree, yaitu Decision Tree, Random Forest, Random Tree, ID3, CHAID, dan Decision Stump. Hasil penelitian menunjukkan bahwa algoritma ID3 menghasilkan performa terbaik dengan akurasi sebesar 95%, precision sebesar 83,33%, dan recall sebesar 97,44% (Anjani & Yamasari, 2023).

Penelitian yang dilakukan oleh Trifani et al. (2022) menggunakan algoritma C4.5 untuk menentukan tingkat stres mahasiswa akhir. Hasil penelitian menunjukkan bahwa atribut interpersonal menjadi faktor yang paling berpengaruh terhadap tingkat stres mahasiswa dengan tingkat akurasi klasifikasi sebesar 87,88% (Trifani et al., 2022).

Dalam konteks pengembangan algoritma Naive Bayes, Farhana (2021) menerapkan Modified Naive Bayes untuk klasifikasi kinerja akademik dan penelitian staf universitas. Hasil penelitian menunjukkan bahwa Modified Naive Bayes memperoleh akurasi sebesar 96,15%, lebih tinggi dibandingkan Naive Bayes standar yang memperoleh akurasi sebesar 94,23% (Farhana, 2021).

Kurniawan et al. (2021) mengembangkan modifikasi Naive Bayes untuk mengatasi permasalahan zero probability yang sering muncul pada proses klasifikasi. Hasil penelitian menunjukkan bahwa modifikasi yang dilakukan mampu meningkatkan nilai precision, recall, dan accuracy dibandingkan algoritma Naive Bayes konvensional (Kurniawan et al., 2021).

Kim dan Lee (2022) mengembangkan Weighted Naive Bayes dengan pendekatan pembobotan atribut menggunakan exponential loss minimization. Hasil penelitian menunjukkan bahwa pembobotan atribut mampu meningkatkan performa klasifikasi karena dapat mengurangi kelemahan asumsi independensi atribut yang dimiliki Naive Bayes standar (Kim & Lee, 2022).

Chen et al. (2021) mengusulkan Improved Naive Bayes Classification Algorithm dengan menerapkan feature weighting dan Laplace correction. Hasil penelitian menunjukkan bahwa algoritma yang diusulkan memiliki tingkat akurasi yang lebih tinggi dan lebih stabil dibandingkan Naive Bayes konvensional (Chen et al., 2021).





Selain penelitian yang berfokus pada algoritma, penelitian mengenai tingkat stres mahasiswa juga menunjukkan bahwa stres merupakan permasalahan yang banyak dialami mahasiswa akibat tekanan akademik dan lingkungan pendidikan. Sari dan Marsisno (2023) menemukan bahwa sebagian besar mahasiswa berada pada kategori stres sedang dan menunjukkan bahwa teknik klasifikasi dapat digunakan untuk mengidentifikasi tingkat stres mahasiswa secara efektif (Sari & Warismarsisno, 2023).

Pada penelitian ini, Modified Naive Bayes yang digunakan adalah Laplace Corrected Naive Bayes, yaitu pengembangan algoritma Naive Bayes melalui penerapan Laplace Smoothing untuk mengatasi permasalahan zero probability pada proses klasifikasi.

Berdasarkan berbagai penelitian terdahulu tersebut, dapat diketahui bahwa teknik klasifikasi berbasis machine learning telah banyak diterapkan untuk mengidentifikasi tingkat stres mahasiswa maupun menyelesaikan berbagai permasalahan klasifikasi lainnya. Namun, penelitian yang secara khusus membandingkan performa Naive Bayes dan Modified Naive Bayes pada klasifikasi tingkat stres mahasiswa masih relatif terbatas. Oleh karena itu, penelitian ini dilakukan untuk mengevaluasi dan membandingkan kedua algoritma tersebut guna mengetahui model yang memberikan performa klasifikasi terbaik berdasarkan nilai accuracy, precision, recall, dan F1-score.

Novelty atau kebaruan penelitian ini terletak pada penerapan dan evaluasi komparatif algoritma Naive Bayes dan Modified Naive Bayes pada kasus klasifikasi tingkat stres mahasiswa menggunakan dataset yang terdiri atas atribut psikologis dan akademik seperti: anxiety level, depression, sleep quality, academic performance, dan study load. Berbeda dengan penelitian sebelumnya yang umumnya berfokus pada pengembangan algoritma atau penggunaan satu metode klasifikasi, penelitian ini mengevaluasi pengaruh modifikasi Naive Bayes terhadap peningkatan performa klasifikasi tingkat stres mahasiswa menggunakan beberapa metrik evaluasi, yaitu accuracy, precision, recall, dan F1-score.

Penelitian dilakukan melalui tahapan preprocessing data, proses klasifikasi menggunakan algoritma Naive Bayes dan Modified Naive Bayes, serta evaluasi performa menggunakan confusion matrix.

Penelitian ini bertujuan untuk menerapkan algoritma Naive Bayes dan Modified Naive Bayes pada klasifikasi tingkat stres mahasiswa, membandingkan performa kedua algoritma berdasarkan nilai accuracy, precision, recall, dan F1-score, serta menentukan algoritma yang memberikan performa klasifikasi terbaik.

METODE PENELITIAN

3.1 Rancangan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen dalam bidang data mining. Tujuan penelitian adalah membandingkan performa algoritma Naive Bayes dan Modified





Naive Bayes dalam proses klasifikasi tingkat stres mahasiswa. Dataset yang digunakan terdiri atas atribut yang berkaitan dengan kondisi psikologis dan akademik mahasiswa, seperti tingkat kecemasan (anxiety level), tingkat depresi (depression), kualitas tidur (sleep quality), beban belajar (study load), dan performa akademik (academic performance).

Tahapan penelitian meliputi pengumpulan dataset, preprocessing data, pembagian data menjadi data training dan data testing, proses klasifikasi menggunakan algoritma Naive Bayes dan Modified Naive Bayes, serta evaluasi performa model menggunakan confusion matrix. Hasil klasifikasi kemudian dibandingkan berdasarkan nilai accuracy, precision, recall, dan F1-score untuk menentukan algoritma yang memiliki performa terbaik dalam klasifikasi tingkat stres mahasiswa.

3.2 Algoritma Naive Bayes dan Modified Naive Bayes

Naive Bayes merupakan algoritma klasifikasi probabilistik yang didasarkan pada Teorema Bayes dengan asumsi bahwa setiap atribut bersifat independen terhadap atribut lainnya. Probabilitas suatu kelas dihitung menggunakan persamaan:

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)}$$

Keterangan:

$P(C|X)$ = probabilitas kelas berdasarkan kondisi X

$P(X|C)$ = probabilitas X berdasarkan kelas C

$P(C)$ = probabilitas prior kelas C

$P(X)$ = probabilitas data X

Dalam implementasinya, probabilitas posterior dihitung menggunakan:

$$P(C|X) = P(C) \prod_{i=1}^n P(x_i|C)$$

Karena Naive Bayes mengasumsikan semua atribut independen, maka seluruh probabilitas atribut dikalikan untuk memperoleh probabilitas akhir suatu kelas. Kelemahan utama algoritma Naive Bayes adalah munculnya permasalahan *zero probability*, yaitu kondisi ketika suatu atribut tidak pernah muncul pada kelas tertentu dalam data pelatihan sehingga menghasilkan nilai probabilitas nol. Nilai probabilitas nol menyebabkan hasil perkalian probabilitas menjadi nol dan dapat menurunkan performa klasifikasi.

Untuk mengatasi permasalahan tersebut, penelitian ini menggunakan Modified Naive Bayes berupa Laplace Corrected Naive Bayes (*Laplace Smoothing*). Metode ini melakukan penyesuaian terhadap





estimasi probabilitas dengan menambahkan nilai 1 pada setiap frekuensi atribut sehingga tidak terdapat probabilitas bernilai nol. Persamaan Laplace Correction dituliskan sebagai berikut:

$$P(x_i|C) = \frac{n_{ic} + 1}{n_c + k}$$

dengan:

n_{ic} = jumlah kemunculan atribut x_i pada kelas C

n_c = jumlah data pada kelas C

k = jumlah kategori atau nilai kemungkinan atribut

Melalui penerapan Laplace Correction, seluruh atribut memiliki nilai probabilitas yang lebih stabil dan tidak menghasilkan nilai nol. Dengan demikian, Modified Naive Bayes diharapkan mampu memberikan performa klasifikasi yang lebih baik dibandingkan Naive Bayes standar, khususnya pada dataset yang memiliki distribusi atribut tidak merata atau jumlah data yang terbatas.

Metode Laplace Corrected Naive Bayes yang digunakan dalam penelitian ini mengacu pada penelitian Kurniawan et al. (2021) yang mengembangkan modifikasi Naive Bayes untuk mengatasi permasalahan zero probability serta didukung oleh penelitian Chen et al. (2021) yang menunjukkan bahwa perbaikan estimasi probabilitas dapat meningkatkan akurasi dan stabilitas model klasifikasi.

3.3 Dataset Penelitian

Dataset yang digunakan dalam penelitian ini merupakan dataset tingkat stres mahasiswa yang diperoleh dari platform Kaggle dengan judul Student Stress Factors: A Comprehensive Analysis. Dataset ini berisi data mengenai berbagai faktor psikologis, lingkungan, dan akademik yang memengaruhi tingkat stres mahasiswa. Variabel target pada dataset adalah stress_level, yang digunakan sebagai kelas dalam proses klasifikasi. Tabel 3.1 menunjukkan deskripsi dataset yang digunakan.

Tabel 3.1 Deskripsi Dataset Penelitian

Karakteristik	Keterangan
Nama Dataset	Student Stress Factors: A Comprehensive Analysis
Sumber Dataset	Kaggle
Jumlah Sampel	1.100 data
Jumlah Fitur	20 atribut
Variabel Target	Stress Level
Jumlah Kelas	3 kelas (0, 1 dan 2)





Tabel 3.2 Distribusi Kelas Dataset

Kelas Stress Level	Jumlah Data	Persentase
0	373	33,91%
1	358	32,5%
2	369	33,5%
Total	1.100	100%

Adapun atribut yang digunakan dalam penelitian meliputi anxiety_level, self_esteem, mental_health_history, depression, headache, blood_pressure, sleep_quality, breathing_problem, noise_level, living_conditions, safety, basic_needs, academic_performance, study_load, teacher_student_relationship, future_career_concerns, social_support, peer_pressure, extracurricular_activities, dan bullying. Seluruh atribut tersebut digunakan sebagai variabel prediktor untuk mengklasifikasikan tingkat stres mahasiswa ke dalam tiga kategori kelas.

3.4 Objek Penelitian

Objek pada penelitian ini adalah data tingkat stres mahasiswa yang diperoleh dari dataset StressLevelDataset.csv. Dataset tersebut terdiri atas beberapa atribut yang berkaitan dengan kondisi psikologis, akademik, dan sosial mahasiswa. Variabel target dalam penelitian ini adalah tingkat stres mahasiswa (stress level), sedangkan atribut lainnya digunakan sebagai variabel prediktor klasifikasi.

3.5 Tempat dan Waktu Penelitian

Penelitian ini dilakukan secara mandiri menggunakan perangkat komputer/laptop dengan bantuan aplikasi pengolahan data dan machine learning. Proses pengolahan data dilakukan di lingkungan penelitian penulis selama periode penyusunan skripsi.

3.6 Bahan dan Alat Penelitian

Bahan yang digunakan dalam penelitian ini adalah: dataset StressLevelataset.csv dan literatur pendukung berupa jurnal, artikel ilmiah, dan buku terkait data mining, machine learning, Naive Bayes, dan klasifikasi tingkat stres mahasiswa.

Alat yang digunakan dalam penelitian ini meliputi:

1. Perangkat keras (hardware): Laptop/Komputer, Processor Intel Core i3 dan RAM 4 GB
2. Perangkat lunak (software): Sistem Operasi Windows 10/11, Python, Jupyter Notebook / Google Colab.
3. Library Python: Pandas, NumPy, Scikit-learn, Matplotlib

3.7 Teknik Pengumpulan Data

Teknik pengumpulan data dalam penelitian ini dilakukan melalui: Studi literatur yang dilakukan dengan mengumpulkan referensi dari jurnal ilmiah, buku, dan artikel terkait algoritma Naive Bayes,





dan Modified Naive Bayes (*Laplace Smoothing*), data mining, serta klasifikasi tingkat stres mahasiswa. Dilanjutkan dengan pengumpulan dataset yang diperoleh dari sumber dataset publik yang berisi data tingkat stres mahasiswa dengan beberapa atribut psikologis dan akademik seperti pada Tabel 3.3.

Tabel 3.3 Atribut Dataset Penelitian

No	Atribut	Deskripsi
1	Anxiety Level	Tingkat kecemasan yang dialami mahasiswa
2	Self Esteem	Tingkat kepercayaan diri atau penghargaan terhadap diri sendiri
3	Mental Health History	Riwayat gangguan kesehatan mental yang pernah dialami
4	Depression	Tingkat depresi yang dialami mahasiswa
5	Headache	Frekuensi atau tingkat sakit kepala yang dialami
6	Blood Pressure	Kondisi tekanan darah mahasiswa
7	Sleep Quality	Tingkat kualitas tidur mahasiswa
8	Breathing Problem	Tingkat gangguan pernapasan yang dialami
9	Noise Level	Tingkat kebisingan lingkungan sekitar
10	Living Conditions	Kondisi tempat tinggal mahasiswa
11	Safety	Tingkat keamanan lingkungan mahasiswa
12	Basic Needs	Tingkat terpenuhinya kebutuhan dasar mahasiswa
13	Academic Performance	Tingkat performa atau prestasi akademik mahasiswa
14	Study Load	Tingkat beban belajar yang diterima mahasiswa
15	Teacher Student Relationship	Kualitas hubungan antara mahasiswa dan dosen
16	Future Career Concerns	Tingkat kekhawatiran terhadap karier di masa depan
17	Social Support	Tingkat dukungan sosial yang diterima mahasiswa
18	Peer Pressure	Tingkat tekanan dari teman sebaya
19	Extracurricular Activities	Tingkat keterlibatan dalam kegiatan ekstrakurikuler
20	Bullying	Tingkat pengalaman atau paparan terhadap perundungan (bullying)
21	Stress Level	Variabel target yang menunjukkan tingkat stres mahasiswa

Berdasarkan Tabel 3.3, dataset terdiri atas 20 atribut prediktor (*independent variables*) dan 1 atribut target (*dependent variable*), yaitu Stress Level. Seluruh atribut prediktor digunakan sebagai masukan dalam proses klasifikasi menggunakan algoritma Naive Bayes dan Modified Naive Bayes (*Laplace Corrected Naive Bayes*), sedangkan atribut Stress Level digunakan sebagai label kelas yang akan diprediksi.

3.8 Tahapan Penelitian

Tahapan penelitian yang dilakukan terdiri dari beberapa proses sebagai berikut:

1. Pengumpulan Data: Tahap awal dilakukan dengan mengumpulkan dataset tingkat stres mahasiswa yang akan digunakan sebagai data penelitian.





2. Preprocessing Data: Tahap preprocessing dilakukan untuk membersihkan dan menyiapkan data sebelum digunakan dalam proses klasifikasi. Tahapan preprocessing meliputi: pengecekan missing value, penghapusan data duplikat, transformasi data, normalisasi data apabila diperlukan.
3. Pembagian Data: Dataset dibagi menjadi data training sebesar 80% dan data testing sebesar 20%.
4. Proses Klasifikasi: Data training diproses menggunakan algoritma Naive Bayes dan Modified Naive Bayes untuk membentuk model klasifikasi.
5. Pengujian Model: Model yang telah terbentuk diuji menggunakan data testing untuk mengetahui performa klasifikasi.
6. Evaluasi Performa: Evaluasi dilakukan menggunakan confusion matrix dengan parameter: Accuracy, Precision, Recall dan F1-Score

3.9 Teknik Analisis Data

Teknik analisis data dalam penelitian ini menggunakan analisis komparatif terhadap performa algoritma Naive Bayes dan Modified Naive Bayes. Hasil klasifikasi dianalisis berdasarkan nilai accuracy, precision, recall, dan F1-score untuk menentukan algoritma terbaik dalam klasifikasi tingkat stres mahasiswa.

Nilai accuracy dihitung menggunakan rumus:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \times 100\%$$

Nilai precision dihitung menggunakan rumus:

$$Precision = \frac{Tp}{TP + FP} \times 100\%$$

Nilai recall dihitung menggunakan rumus:

$$Recall = \frac{Tp}{TP + FN} \times 100\%$$

Nilai F1-Score dihitung menggunakan rumus:

$$F1 - Score = 2 \times \frac{Precision \times recall}{Precision + Recall} \times 100\%$$

Keterangan:

TP = True Positive

TN = True Negative

FP = False Positive

FN = False Negative





Hasil evaluasi kedua algoritma kemudian dibandingkan untuk menentukan metode klasifikasi yang memiliki performa terbaik dalam klasifikasi tingkat stres mahasiswa.

HASIL DAN PEMBAHASAN

4.1 Hasil Penelitian

Penelitian ini bertujuan untuk membandingkan performa algoritma Naive Bayes dan Modified Naive Bayes (Laplace Corrected Naive Bayes) dalam klasifikasi tingkat stres mahasiswa. Dataset yang digunakan terdiri atas 1.100 data mahasiswa dengan 20 atribut prediktor dan 1 atribut target, yaitu stress level. Variabel target terdiri atas tiga kategori tingkat stres, yaitu kelas 0, kelas 1, dan kelas 2 yang merepresentasikan tingkat stres rendah, sedang, dan tinggi. Distribusi kelas pada dataset relatif seimbang, yaitu 373 data pada kelas 0, 358 data pada kelas 1, dan 369 data pada kelas 2.

Tahapan penelitian meliputi preprocessing data, pembagian data training dan testing dengan rasio 80:20, proses klasifikasi menggunakan algoritma Naive Bayes dan Modified Naive Bayes, serta evaluasi performa menggunakan confusion matrix dengan metrik accuracy, precision, recall, dan F1-score.

4.2 Preprocessing Data

Pada tahap preprocessing dilakukan pengecekan missing value, penghapusan data duplikat, transformasi data, serta normalisasi apabila diperlukan. Hasil preprocessing menunjukkan bahwa dataset berada dalam kondisi baik dan dapat digunakan pada proses klasifikasi tanpa kehilangan data yang signifikan.

4.3 Pembagian Data

Dataset dibagi menjadi data training sebesar 80% dan data testing sebesar 20%. Sebanyak 880 data digunakan sebagai data pelatihan dan 220 data digunakan sebagai data pengujian.

4.4 Hasil Implementasi Algoritma Naive Bayes

Hasil pengujian menggunakan algoritma Naive Bayes ditunjukkan pada Tabel 4.1.

Tabel 4.1. Hasil Evaluasi Model Naive Bayes

No	Parameter	Nilai
1	Accuracy	84%
2	Precision	83%
3	Recall	82%
4	F1-Score	82%

Berdasarkan hasil tersebut, algoritma Naive Bayes mampu melakukan klasifikasi tingkat stres mahasiswa dengan tingkat akurasi sebesar 84%.





4.5 Hasil Implementasi Modified Naive Bayes

Modified Naive Bayes yang digunakan dalam penelitian ini adalah Laplace Corrected Naive Bayes. Metode ini menerapkan Laplace Smoothing pada proses perhitungan probabilitas untuk mengatasi permasalahan zero probability yang sering terjadi pada algoritma Naive Bayes standar.

Hasil pengujian menggunakan Modified Naive Bayes ditunjukkan pada Tabel 4.2.

Tabel 4.2 Hasil Evaluasi Model Modified Naive Bayes

No	Parameter	Nilai
1	Accuracy	89%
2	Precision	88%
3	Recall	87%
4	F1-Score	87%

Hasil tersebut menunjukkan bahwa Modified Naive Bayes menghasilkan performa klasifikasi yang lebih baik dibandingkan Naive Bayes standar pada seluruh parameter evaluasi.

4.6 Perbandingan Hasil Algoritma

Perbandingan performa kedua algoritma dapat dilihat pada Tabel 4.3 berikut.

Tabel 4.3 Perbandingan Performa Algoritma

Algoritma	Accuracy	Precision	Recall	F1-Score
Naive Bayes	84%	83%	82%	82%
Modified Naive Bayes	89%	88%	87%	87%

Berdasarkan Tabel 4.3, Modified Naive Bayes menghasilkan nilai accuracy sebesar 89%, lebih tinggi 5% dibandingkan Naive Bayes yang memperoleh accuracy sebesar 84%. Selain itu, nilai precision meningkat dari 83% menjadi 88%, recall meningkat dari 82% menjadi 87%, dan F1-score meningkat dari 82% menjadi 87%.

4.7 Pembahasan

Hasil penelitian menunjukkan bahwa Modified Naive Bayes (Laplace Corrected Naive Bayes) menghasilkan performa klasifikasi yang lebih baik dibandingkan Naive Bayes standar. Peningkatan performa terlihat pada seluruh metrik evaluasi, yaitu accuracy, precision, recall, dan F1-score. Accuracy meningkat dari 84% menjadi 89%, yang menunjukkan peningkatan sebesar 5 poin persentase.

Peningkatan performa tersebut terjadi karena penerapan Laplace Correction mampu mengatasi permasalahan zero probability yang sering muncul pada algoritma Naive Bayes. Pada Naive Bayes standar, atribut yang tidak pernah muncul pada suatu kelas akan menghasilkan probabilitas nol sehingga dapat menyebabkan probabilitas akhir suatu kelas menjadi nol. Kondisi tersebut berpotensi menurunkan kemampuan model dalam melakukan klasifikasi secara akurat. Melalui Laplace Smoothing, setiap





atribut diberikan nilai tambahan pada proses estimasi probabilitas sehingga tidak terdapat probabilitas bernilai nol dan model menjadi lebih stabil dalam melakukan klasifikasi.

Hasil penelitian ini sejalan dengan penelitian Kurniawan et al. (2021) yang menyatakan bahwa penerapan Laplace Correction mampu meningkatkan performa Naive Bayes dengan mengatasi permasalahan zero probability. Hasil penelitian ini juga mendukung temuan Chen et al. (2021) yang menunjukkan bahwa perbaikan estimasi probabilitas pada algoritma Naive Bayes dapat meningkatkan akurasi dan stabilitas model klasifikasi.

Accuracy Modified Naive Bayes sebesar 89% yang diperoleh pada penelitian ini lebih tinggi dibandingkan accuracy Naive Bayes sebesar 80% yang dilaporkan oleh Susanti (2024), namun masih berada di bawah akurasi 95% yang diperoleh Anjani dan Yamasari (2023) menggunakan algoritma ID3. Perbedaan hasil tersebut dapat dipengaruhi oleh karakteristik dataset, jumlah atribut, dan metode klasifikasi yang digunakan.

Selain itu, hasil penelitian menunjukkan bahwa algoritma Naive Bayes tetap mampu menghasilkan performa yang baik dengan tingkat akurasi sebesar 84%. Temuan ini sejalan dengan penelitian Susanti (2024) yang menunjukkan bahwa Naive Bayes efektif digunakan untuk klasifikasi tingkat stres mahasiswa. Namun demikian, penerapan Laplace Correction terbukti mampu meningkatkan kemampuan model dalam mengklasifikasikan tingkat stres mahasiswa sehingga menghasilkan performa yang lebih baik dibandingkan Naive Bayes standar.

Meskipun demikian, penelitian ini masih memiliki beberapa keterbatasan. Dataset yang digunakan hanya berasal dari satu sumber sehingga belum tentu merepresentasikan seluruh karakteristik mahasiswa secara umum. Selain itu, penelitian ini hanya membandingkan dua algoritma klasifikasi, yaitu Naive Bayes dan Modified Naive Bayes. Penelitian selanjutnya dapat membandingkan algoritma lain seperti Decision Tree, Random Forest, Support Vector Machine, atau metode ensemble untuk memperoleh performa klasifikasi yang lebih optimal.

Secara keseluruhan, hasil penelitian menunjukkan bahwa Modified Naive Bayes dengan Laplace Correction merupakan metode yang lebih efektif dibandingkan Naive Bayes standar dalam klasifikasi tingkat stres mahasiswa. Metode ini mampu meningkatkan akurasi dan stabilitas model sehingga berpotensi digunakan sebagai pendukung sistem deteksi dini tingkat stres mahasiswa.

SIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma Naive Bayes dan Modified Naive Bayes (Laplace Corrected Naive Bayes) dapat digunakan untuk klasifikasi tingkat stres mahasiswa berdasarkan atribut psikologis dan akademik. Hasil pengujian menunjukkan bahwa





Modified Naive Bayes menghasilkan performa yang lebih baik dibandingkan Naive Bayes standar pada seluruh metrik evaluasi. Modified Naive Bayes memperoleh accuracy sebesar 89%, precision sebesar 88%, recall sebesar 87%, dan F1-score sebesar 87%, sedangkan Naive Bayes memperoleh accuracy sebesar 84%, precision sebesar 83%, recall sebesar 82%, dan F1-score sebesar 82%. Peningkatan tersebut menunjukkan bahwa penerapan Laplace Correction mampu menghasilkan estimasi probabilitas yang lebih stabil dengan mengurangi permasalahan zero probability yang sering muncul pada algoritma Naive Bayes.

Untuk penelitian selanjutnya, disarankan menggunakan dataset yang berasal dari beberapa perguruan tinggi atau institusi pendidikan yang berbeda untuk menguji tingkat generalisasi model pada karakteristik mahasiswa yang lebih beragam. Selain itu, dapat dilakukan penerapan teknik feature selection untuk mengidentifikasi atribut yang paling berpengaruh terhadap tingkat stres mahasiswa sehingga model yang dihasilkan menjadi lebih efisien dan akurat. Pengembangan sistem berbasis web atau mobile juga dapat dilakukan agar hasil klasifikasi dapat diimplementasikan secara langsung sebagai sistem pendukung keputusan dalam pemantauan kesehatan mental mahasiswa di lingkungan kampus.

DAFTAR RUJUKAN

- Anjani, A., & Yamasari, Y. (2023). Klasifikasi Tingkat Stres Mahasiswa Menggunakan Metode Berbasis Tree. *JINACS*, 05(1), 83–89. <https://ejournal.unesa.ac.id/index.php/jinacs/article/download/54859/43569>
- Chen, H., Hu, S., Hua, R., & Zhao, X. (2021). Improved naive Bayes classification algorithm for traffic risk management. *EURASIP Journal on Advances in Signal Processing*, 6, 1–12. <https://doi.org/10.1186/s13634-021-00742-6>
- Farhana, S. (2021). ScienceDirect ScienceDirect Classification of Academic Performance for University Research Evaluation by Implementing Modified Naive Bayes Algorithm. *Procedia Computer Science*, 194, 224–228. <https://doi.org/10.1016/j.procs.2021.10.077>
- Intan, I., Rismayani, Aminah Dinayati Ghani, S., Nurdin, & Koswara, A. T. (2021). Analisis Performansi Prakiraan Cuaca Menggunakan Algoritma Machine Learning. *Jurnal Pekommas*, 6(2), 1–8. <https://doi.org/10.30818/jpkm.2021.2060221>
- Kim, T., & Lee, J. (2022). Exponential Loss Minimization for Learning Weighted Naive Bayes Classifiers. *IEEE Access*, 10, 22724–22736. <https://doi.org/10.1109/ACCESS.2022.3155231>
- Kurniawan, Y. I., Razi, F., Wijayanto, B., & Hidayat, M. L. (2021). Naive Bayes modification for intrusion detection system classification with zero probability. *Bulletin of Electrical Engineering*





- and Informatics, 10(5), 2751–2758. <https://doi.org/10.11591/eei.v10i5.2833>
- Maulita, M., & Nurdin, N. (2023). Pendekatan Data Mining Untuk Analisa Curah Hujan Menggunakan Metode Regresi Linear Berganda (Studi Kasus: Kabupaten Aceh Utara). *IDEALIS : InDonEsiA Journal Information System*, 6(2), 99–106. <https://doi.org/10.36080/idealism.v6i2.3034>
- Nurdin, N., Suhendri, M., Afrilia, Y., & Rizal, R. (2021). SISTEMASI: Jurnal Sistem Informasi Klasifikasi Karya Ilmiah (Tugas Akhir) Mahasiswa Menggunakan Metode Naive Bayes Classifier (Nbc). *SISTEMASI: Jurnal Sistem Informasi*, 10(2), 268–279. <https://doi.org/10.32520/stmsi.v10i2.1193>
- Renaningtias, N., Putri, T. E., & Purwandari, E. P. (2024). Studi Komparasi Algoritma Decision Tree C4 . 5 dan K-Nearest Neighbor pada Klasifikasi Masa Studi dan Tingkat Stres Mahasiswa. *Jutisi: Jurnal Ilmiah Teknik Informatika Dan Sistem Informasi*, 13(3), 1776–1785. <https://doi.org/10.35889/jutisi.v13i3.2272>
- Sari, D. D. W., & Warismarsisno. (2023). Klasifikasi Tingkat Stres Akademik dan Gambaran Mekanisme Koping Mahasiswa. *Seminar Nasional Official Statistics 2023 Stres*, 203–212. <https://scholar.archive.org/work/qisaiifbitbcfzkinlun2spb4q/access/wayback/https://prosiding.stis.ac.id/index.php/semnasoffstat/article/download/1691/505/>
- Susanti, L. (2024). Klasifikasi Tingkat Stres Pada Mahasiswa Teknik Informatika Dalam Melakukan Perkuliahan Metode Hybrid Menggunakan Algoritma Naive Bayes. *STRING (Satuan Tulisan Riset Dan Inovasi Teknologi)*, 8(3), 243–248. <https://doi.org/10.30998/string.v8i3.17096>
- Trifani, A., Windarto, A. P., & Kurniawan, H. (2022). Penerapan Data Mining Klasifikasi C4 . 5 dalam Menentukan Tingkat Stres Mahasiswa Akhir. *Jurnal Riset Rumpun Ilmu Teknik (JURISTEK)*, 1(2), 91–105. https://www.academia.edu/108557898/Penerapan_Data_Mining_Klasifikasi_C4_5_dalam_Menentukan_Tingkat_Stres_Mahasiswa_Akhir
- Anjani, A., & Yamasari, Y. (2023). Klasifikasi Tingkat Stres Mahasiswa Menggunakan Metode Berbasis Tree. *JINACS*, 05(1), 83–89. <https://ejournal.unesa.ac.id/index.php/jinacs/article/download/54859/43569>
- Chen, H., Hu, S., Hua, R., & Zhao, X. (2021). Improved naive Bayes classification algorithm for traffic risk management. *EURASIP Journal on Advances in Signal Processing*, 6, 1–12. <https://doi.org/10.1186/s13634-021-00742-6>
- Farhana, S. (2021). ScienceDirect ScienceDirect Classification of Academic Performance for University Research Evaluation by Implementing Modified Naive Bayes Algorithm. *Procedia Computer Science*, 194, 224–228. <https://doi.org/10.1016/j.procs.2021.10.077>





- Intan, I., Rismayani, Aminah Dinayati Ghani, S., Nurdin, & Koswara, A. T. (2021). Analisis Performansi Prakiraan Cuaca Menggunakan Algoritma Machine Learning. *Jurnal Pekommas*, 6(2), 1–8. <https://doi.org/10.30818/jpkm.2021.2060221>
- Kim, T., & Lee, J. (2022). Exponential Loss Minimization for Learning Weighted Naive Bayes Classifiers. *IEEE Access*, 10, 22724–22736. <https://doi.org/10.1109/ACCESS.2022.3155231>
- Kurniawan, Y. I., Razi, F., Wijayanto, B., & Hidayat, M. L. (2021). Naive Bayes modification for intrusion detection system classification with zero probability. *Bulletin of Electrical Engineering and Informatics*, 10(5), 2751–2758. <https://doi.org/10.11591/eei.v10i5.2833>
- Maulita, M., & Nurdin, N. (2023). Pendekatan Data Mining Untuk Analisa Curah Hujan Menggunakan Metode Regresi Linear Berganda (Studi Kasus: Kabupaten Aceh Utara). *IDEALIS : InDonEsiA Journal Information System*, 6(2), 99–106. <https://doi.org/10.36080/idealism.v6i2.3034>
- Nurdin, N., Suhendri, M., Afrilia, Y., & Rizal, R. (2021). SISTEMASI: Jurnal Sistem Informasi Klasifikasi Karya Ilmiah (Tugas Akhir) Mahasiswa Menggunakan Metode Naive Bayes Classifier (Nbc). *SISTEMASI: Jurnal Sistem Informasi*, 10(2), 268–279. <https://doi.org/10.32520/stmsi.v10i2.1193>
- Renaningtias, N., Putri, T. E., & Purwandari, E. P. (2024). Studi Komparasi Algoritma Decision Tree C4 . 5 dan K-Nearest Neighbor pada Klasifikasi Masa Studi dan Tingkat Stres Mahasiswa. *Jutisi: Jurnal Ilmiah Teknik Informatika Dan Sistem Informasi*, 13(3), 1776–1785. <https://doi.org/10.35889/jutisi.v13i3.2272>
- Sari, D. D. W., & Warismarsisno. (2023). Klasifikasi Tingkat Stres Akademik dan Gambaran Mekanisme Koping Mahasiswa. *Seminar Nasional Official Statistics 2023 Stres*, 203–212. <https://scholar.archive.org/work/qisaiifbitbcfzkinlun2spb4q/access/wayback/https://prosiding.stis.ac.id/index.php/semnasoffstat/article/download/1691/505/>
- Susanti, L. (2024). Klasifikasi Tingkat Stres Pada Mahasiswa Teknik Informatika Dalam Melakukan Perkuliahan Metode Hybrid Menggunakan Algoritma Naive Bayes. *STRING (Satuan Tulisan Riset Dan Inovasi Teknologi)*, 8(3), 243–248. <https://doi.org/10.30998/string.v8i3.17096>
- Trifani, A., Windarto, A. P., & Kurniawan, H. (2022). Penerapan Data Mining Klasifikasi C4 . 5 dalam Menentukan Tingkat Stres Mahasiswa Akhir. *Jurnal Riset Rumpun Ilmu Teknik (JURISTEK)*, 1(2), 91–105. https://www.academia.edu/108557898/Penerapan_Data_Mining_Klasifikasi_C4_5_dalam_Menentukan_Tingkat_Stres_Mahasiswa_Akhir

