

Komparasi Algoritma Machine Learning untuk Prediksi Risiko Burnout Akademis Akibat Penggunaan Generative AI

Ophelius Savero Rinyuakng Rolin^{1*}, Wahyu Nugraha², Rabiatus Sa'adah³

^{1,2,3}Fakultas Teknik dan Informatika, Informatika, Universitas Bina Sarana Informatika, Pontianak, Indonesia

Email: ^{1*}riukmd@gmail.com, ²wahyu.wnu@bsi.ac.id, ³rabiatus.rbs@bsi.ac.id

(*Email Corresponding Author: riukmd@gmail.com)

Received: 28 Juni 2026 | Revision: 4 Juli 2026 | Accepted: 7 Juli 2026

Abstrak

Perkembangan Generative AI (GenAI) dalam pendidikan tinggi memunculkan fenomena baru berupa peningkatan risiko burnout akademis mahasiswa yang tidak lagi semata dipicu oleh beban belajar konvensional, melainkan oleh pola interaksi dengan kecerdasan buatan. Penelitian ini bertujuan membandingkan kinerja tiga algoritma Machine Learning, yaitu Logistic Regression, Random Forest Classifier, dan XGBoost, dalam mengklasifikasikan tingkat risiko burnout akademis (High, Medium, Low) berdasarkan data profil 50.000 mahasiswa dengan 16 atribut, termasuk durasi penggunaan GenAI mingguan (Weekly_GenAI_Hours) dan tingkat ketergantungan psikologis terhadap AI (Perceived_AI_Dependency). Metode penelitian meliputi eksplorasi data, pra-pemrosesan (encoding, normalisasi, pembagian data latih-uji 80:20), pelatihan model baseline, optimasi hyperparameter menggunakan GridSearchCV, serta evaluasi menggunakan metrik akurasi, presisi, recall, F1-score, confusion matrix, dan feature importance. Hasil penelitian menunjukkan akurasi model berkisar antara 51,96% hingga 53,83%, dengan Logistic Regression baseline mencatatkan akurasi tertinggi (53,83%), diikuti XGBoost setelah tuning (53,62%) dan Random Forest setelah tuning (53,27%). Hyperparameter tuning terbukti meningkatkan akurasi model berbasis pohon secara signifikan, namun tidak pada model linear. Ekstraksi feature importance dari Random Forest dan XGBoost secara konsisten menempatkan Weekly_GenAI_Hours dan Perceived_AI_Dependency sebagai dua prediktor terkuat, jauh melampaui variabel akademik tradisional seperti IPK dan jam belajar. Analisis confusion matrix mengungkap tumpang tindih klasifikasi terutama pada kelas Medium. Temuan ini memberikan bukti empiris bahwa intensitas dan ketergantungan penggunaan GenAI merupakan determinan utama burnout akademis, sehingga dapat menjadi dasar objektif bagi kebijakan pembatasan penggunaan AI yang sehat di lingkungan pendidikan.

Kata Kunci: Burnout Akademis, Generative AI, Machine Learning, Random Forest, XGBoost, Feature Importance

Abstract

The rapid adoption of Generative AI (GenAI) in higher education has introduced a new phenomenon in which academic burnout risk is increasingly driven by patterns of AI interaction rather than conventional study load alone. This study compares the performance of three Machine Learning algorithms, namely Logistic Regression, Random Forest Classifier, and XGBoost, in classifying academic burnout risk levels (High, Medium, Low) using a profile dataset of 50,000 students with 16 attributes, including weekly GenAI usage hours (Weekly_GenAI_Hours) and perceived psychological dependency on AI (Perceived_AI_Dependency). The research method covers exploratory data analysis, preprocessing (encoding, normalization, an 80:20 train-test split), baseline model training, hyperparameter optimization through GridSearchCV, and evaluation using accuracy, precision, recall, F1-score, confusion matrix, and feature importance metrics. The results show model accuracy ranging from 51.96% to 53.83%, with the baseline Logistic Regression achieving the highest accuracy (53.83%), followed by tuned XGBoost (53.62%) and tuned Random Forest (53.27%). Hyperparameter tuning significantly improved the accuracy of tree-based models but slightly reduced that of the linear model. Feature importance extraction from Random Forest and XGBoost consistently ranked Weekly_GenAI_Hours and Perceived_AI_Dependency as the two strongest predictors, far outweighing conventional academic variables such as GPA and study hours. Confusion matrix analysis revealed classification overlap, particularly for the Medium class. These findings provide empirical evidence that the intensity and dependency of GenAI usage are the primary determinants of academic burnout, offering an objective basis for institutional policies on healthy AI usage limits in education.

Keywords: Academic Burnout, Generative AI, Machine Learning, Random Forest, XGBoost, Feature Importance

1. PENDAHULUAN

Perkembangan pesat teknologi kecerdasan buatan generatif (Generative AI/GenAI) telah membawa transformasi signifikan dalam sektor pendidikan tinggi. Kehadiran GenAI mengubah cara mahasiswa berinteraksi dengan informasi, menyelesaikan tugas, dan melakukan pembelajaran mandiri, mulai dari pencarian ide (ideation), peringkasan bahan bacaan, hingga penyusunan jawaban secara instan. Di balik

efisiensi waktu dan kemudahan akses informasi yang ditawarkan, muncul fenomena baru yang berkaitan dengan kesejahteraan mental mahasiswa, yaitu meningkatnya risiko kelelahan akademis atau burnout.

Secara tradisional, burnout akademis dikaitkan dengan beban studi konvensional, seperti lamanya jam belajar (traditional study hours) atau tekanan mempertahankan Indeks Prestasi Kumulatif (IPK) yang tinggi. Namun, pergeseran paradigma belajar yang melibatkan instrumen GenAI membawa dimensi tekanan baru. Tinjauan sistematis terhadap literatur pendidikan menunjukkan bahwa integrasi GenAI di ruang kelas berpotensi memengaruhi kesejahteraan mental mahasiswa melalui pola ketergantungan dan reduksi interaksi sosial [8], sementara studi lain pada mahasiswa pengguna ChatGPT menemukan bahwa intensitas penggunaan tersebut berasosiasi negatif dengan kesejahteraan akademis melalui mekanisme penurunan self-control [9]. Ketergantungan pada GenAI juga terbukti memicu kelelahan kognitif (cognitive fatigue) yang pada akhirnya melemahkan kemampuan berpikir kritis mahasiswa [13]. Eksperimen awal terhadap data profil akademik 50.000 mahasiswa menunjukkan bahwa metrik akademik tradisional bukan lagi faktor penentu utama stres; durasi penggunaan GenAI per minggu (Weekly_GenAI_Hours) justru muncul sebagai prediktor paling signifikan dalam memetakan tingkat risiko burnout. Selain durasi penggunaan, tingkat ketergantungan psikologis mahasiswa terhadap AI (Perceived_AI_Dependency) menempati urutan kedua sebagai prediktor terkuat, mengindikasikan adanya pergeseran sumber tekanan dari intensitas belajar manual menuju kelelahan mental akibat tingginya paparan dan ketergantungan psikologis pada kecerdasan buatan.

Memprediksi profil risiko burnout dengan variabel interaksi AI modern ini merupakan masalah klasifikasi yang kompleks. Pengujian awal menggunakan algoritma Machine Learning seperti Logistic Regression, Random Forest Classifier, dan XGBoost mencatatkan rentang akurasi prediksi antara 51,96% hingga 53,8%. Akurasi yang relatif konstan pada kisaran tersebut mencerminkan tingginya noise dan varians kompleks dari faktor psikologis manusia. Meskipun demikian, kapabilitas algoritma berbasis pohon (tree-based ensemble) seperti Random Forest dan XGBoost tetap krusial untuk dikaji, terutama kemampuannya mengekstraksi feature importance guna mengidentifikasi variabel paling berpengaruh. Kondisi ini diperparah oleh keragaman kebijakan institusional terhadap penggunaan AI, mulai dari pelarangan ketat (strict ban) hingga izin dengan batasan sitasi (allowed with citation), yang mencerminkan ketidakpastian yang masih dihadapi dunia pendidikan dalam merespons adopsi GenAI.

Beberapa penelitian terdahulu telah mengeksplorasi potensi Machine Learning dalam mendeteksi burnout pada populasi pelajar. Yeskuatov dkk. [1] menerapkan pendekatan supervised learning untuk mendeteksi burnout pada mahasiswa ilmu komputer dan menunjukkan bahwa algoritma berbasis pohon mampu mengidentifikasi pola burnout dengan akurasi lebih tinggi dibandingkan metode survei konvensional, namun penelitian tersebut belum menyertakan variabel interaksi GenAI. Pereira dkk. [2] membandingkan Random Forest, XGBoost, dan Support Vector Machine untuk mengklasifikasikan dua profil risiko (Burnout Risk dan No Risk) pada mahasiswa psikologi, dengan akurasi Random Forest mencapai 95,06% dan XGBoost 93,82%; penelitian ini bersifat klasifikasi biner dan tidak melibatkan variabel penggunaan GenAI. Feher dkk. [3] menggunakan Random Forest dengan pembobotan kelas untuk memprediksi burnout guru dan menemukan AUC sebesar 0,811, dengan penggunaan internet berlebihan sebagai prediktor signifikan, tetapi variabel yang digunakan bersifat umum (penggunaan internet), bukan penggunaan GenAI secara spesifik. Sharma dkk. [4] mengembangkan model hibrid berbasis data log Learning Management System untuk deteksi dini burnout mahasiswa dan menegaskan keunggulan pendekatan ensemble berbasis pohon dalam menangani data perilaku kompleks, namun data yang digunakan belum mencakup penggunaan alat GenAI eksternal seperti ChatGPT, Google Gemini, atau Microsoft Copilot yang kini menjadi bagian integral aktivitas akademis mahasiswa.

Kajian yang lebih baru memperluas cakupan prediktor burnout berbasis Machine Learning. Kuadey dkk. [5] menerapkan algoritma Machine Learning untuk mengkaji dampak technostress creators terhadap burnout belajar dan performa akademik mahasiswa, namun belum secara spesifik mengisolasi variabel intensitas penggunaan GenAI. Nellore [6] membandingkan Logistic Regression, Random Forest, dan XGBoost dalam kerangka analitik data untuk memprediksi stres akademik, dengan XGBoost mencatatkan akurasi tertinggi (88,3%), sedangkan Tapio [7] menggunakan model Decision Tree berbasis data psikometrik untuk memprediksi burnout mahasiswa di perguruan tinggi swasta dengan performa moderat. Ketiga penelitian tersebut memperkuat relevansi algoritma ensemble berbasis pohon dalam permasalahan prediksi burnout, namun tetap belum menempatkan variabel interaksi GenAI sebagai fokus prediktor utama.

Berdasarkan tinjauan tersebut, terdapat kesenjangan penelitian (research gap) yang signifikan, yaitu belum adanya kajian yang secara spesifik menempatkan variabel interaksi GenAI durasi penggunaan mingguan dan tingkat ketergantungan psikologis sebagai prediktor utama risiko burnout akademis dalam skema klasifikasi tiga kelas (High, Medium, Low). Penelitian ini berupaya mengisi kesenjangan tersebut

melalui tiga rumusan masalah, yaitu: (1) bagaimana pengaruh pola interaksi kecerdasan buatan terhadap tingkat risiko burnout mahasiswa; (2) bagaimana perbandingan kinerja Logistic Regression, Random Forest Classifier, dan XGBoost dalam memprediksi risiko burnout; dan (3) variabel apa saja yang menjadi prediktor paling signifikan serta bagaimana hasil ekstraksi fitur tersebut dapat digunakan untuk merekomendasikan ambang batas penggunaan GenAI yang sehat. Tujuan penelitian ini adalah menganalisis pengaruh interaksi GenAI terhadap klasifikasi risiko burnout, mengevaluasi dan membandingkan kinerja ketiga algoritma tersebut baik sebelum maupun sesudah hyperparameter tuning menggunakan GridSearchCV, serta mengidentifikasi prediktor utama pemicu stres akademis melalui ekstraksi feature importance sebagai dasar objektif perumusan kebijakan pendidikan terkait penggunaan Generative AI yang sehat.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif yang mengalir secara berurutan, dimulai dari pengumpulan data, eksplorasi data secara visual (Exploratory Data Analysis), pra-pemrosesan data, pelatihan model klasifikasi baseline, optimasi hyperparameter, hingga evaluasi kinerja model dan penarikan simpulan, sebagaimana diringkas pada Gambar 1.



Gambar 1. Tahapan Penelitian

Data mentah yang telah dikumpulkan dieksplorasi secara visual untuk memahami pola dan distribusi antarvariabel. Selanjutnya, data melalui tahap pra-pemrosesan (pembersihan data, encoding variabel kategorikal, normalisasi, dan pembagian data latih-uji) agar siap digunakan oleh algoritma. Model klasifikasi dilatih menggunakan parameter dasar (baseline), kemudian dioptimalkan melalui hyperparameter tuning menggunakan GridSearchCV. Kinerja model akhir dievaluasi menggunakan metrik kuantitatif dan dianalisis lebih lanjut untuk menjawab tujuan penelitian.

2.2 Dataset

Dataset yang digunakan adalah dataset sintesis (synthetic dataset) "Student Burnout Prediction from AI Usage" yang merepresentasikan profil akademis 50.000 mahasiswa dengan 16 atribut, dengan variabel target Burnout_Risk_Level yang terbagi menjadi tiga kelas, yaitu High, Medium, dan Low. Beberapa variabel utama yang digunakan dalam penelitian disajikan pada Tabel 1.

Tabel 1. Deskripsi Variabel Utama Dataset

Nama Variabel	Tipe	Deskripsi
Weekly_GenAI_Hours	Numerik	Durasi penggunaan Generative AI per minggu (jam)
Perceived_AI_Dependency	Ordinal	Tingkat ketergantungan psikologis terhadap AI (skala 1-10)
Pre_Semester_GPA	Numerik	IPK mahasiswa sebelum semester berjalan
Post_Semester_GPA	Numerik	IPK mahasiswa setelah semester berjalan
Traditional_Study_Hours	Numerik	Jumlah jam belajar konvensional per minggu
Anxiety_Level_During_Exams	Ordinal	Tingkat kecemasan saat ujian (skala 1-10)
Skill_Retention_Score	Numerik	Skor retensi keterampilan kognitif mahasiswa
Major_Category	Kategorikal	Kategori program studi/jurusan mahasiswa
Year_of_Study	Ordinal	Tahun studi mahasiswa (1-4)
Institution_AI_Policy	Kategorikal	Kebijakan institusi terhadap penggunaan AI
Burnout_Risk_Level	Kategorikal	Label target: High, Medium, Low

2.3 Pra-pemrosesan Data

Tahap pra-pemrosesan data dilakukan melalui lima teknik utama menggunakan pustaka scikit-learn [14]. Pertama, kolom Student_ID dihapus karena hanya berfungsi sebagai identitas unik tanpa nilai prediktif. Kedua, variabel target Burnout_Risk_Level ditransformasi menjadi nilai numerik menggunakan LabelEncoder. Ketiga, variabel kategorikal lainnya diubah menjadi data biner menggunakan one-hot encoding dengan opsi drop_first=True untuk menghindari

multikolinearitas. Keempat, dataset dibagi menjadi data latih (80%) dan data uji (20%) menggunakan `train_test_split`. Kelima, seluruh fitur numerik dinormalisasi menggunakan `StandardScaler` agar memiliki skala yang seragam sebelum digunakan untuk pelatihan model.

2.4 Desain Model dan Optimasi Hyperparameter

Penelitian ini membandingkan tiga algoritma klasifikasi. Logistic Regression digunakan sebagai model dasar (baseline) dengan parameter default dari pustaka `scikit-learn`. Random Forest Classifier [11] dilatih dengan parameter awal `n_estimators=100` untuk menangkap hubungan non-linear antarfitur melalui pendekatan ensemble berbasis pohon keputusan. XGBoost [10] dilatih dengan `objective multi:softmax` untuk menangani klasifikasi tiga kelas melalui pendekatan gradient boosting yang membangun pohon secara bertahap guna meminimalkan error. Ketiga model kemudian dioptimalkan menggunakan `GridSearchCV` dengan validasi silang, dengan kombinasi parameter yang diujikan disajikan pada Tabel 2.

Tabel 2. Skenario Parameter Hyperparameter Tuning (`GridSearchCV`)

Algoritma	Parameter	Nilai yang Diuji (Grid)
Logistic Regression	C	0,001; 0,01; 0,1; 1; 10; 100
	solver	liblinear; lbfgs; saga
	penalty	l1; l2
Random Forest	n_estimators	50; 100
	max_depth	10; 20
	min_samples_split	2; 5
XGBoost	n_estimators	100; 200
	max_depth	3; 5
	learning_rate	0,1; 0,2
	subsample	0,7; 0,9
	colsample_bytree	0,7; 0,9

2.5 Metrik Evaluasi Kinerja

Kinerja model dievaluasi pada data uji (10.000 sampel) menggunakan empat metrik utama klasifikasi multikelas, yaitu Akurasi, Presisi (weighted), Recall (weighted), dan F1-Score [15], mengikuti formulasi berikut, dengan TP, TN, FP, dan FN masing-masing merepresentasikan True Positive, True Negative, False Positive, dan False Negative:

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \quad (1)$$

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \quad (2)$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) \quad (3)$$

$$\text{F1-Score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (4)$$

Selain keempat metrik tersebut, penelitian ini juga menggunakan Confusion Matrix untuk memvisualisasikan distribusi hasil prediksi pada masing-masing kelas, serta ekstraksi Feature Importance dari algoritma berbasis pohon (Random Forest dan XGBoost) untuk mengidentifikasi variabel paling berpengaruh terhadap prediksi model.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Eksplorasi Data (EDA)

Analisis data eksploratif dilakukan terhadap 50.000 data observasi untuk memahami pola distribusi variabel sebelum pemodelan. Distribusi variabel target `Burnout_Risk_Level` menunjukkan ketidakseimbangan kelas (imbalanced), sebagaimana disajikan pada Tabel 3.

Tabel 3. Distribusi Kelas `Burnout_Risk_Level`

Kelas	Jumlah Sampel	Proporsi (%)
High	12.487	24,97
Low	16.369	32,74
Medium	21.144	42,29
Total	50.000	100,00

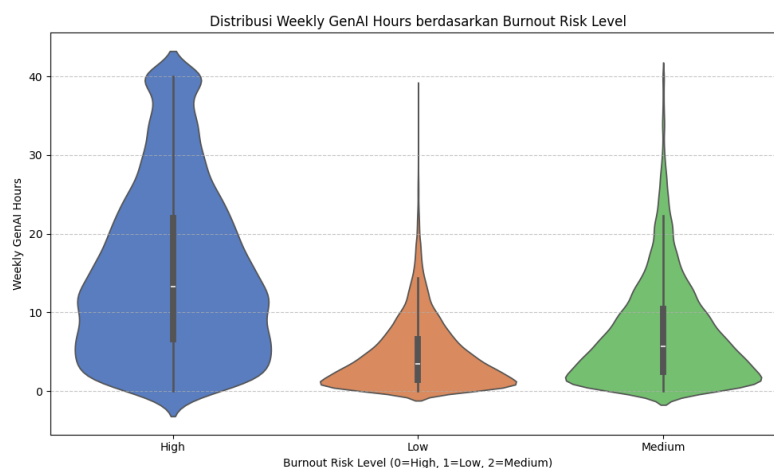
Kelas Medium merupakan kelas dengan proporsi terbesar (42,29%), diikuti kelas Low (32,74%), sedangkan kelas High menjadi kelas minoritas (24,97%). Ketidakseimbangan ini menjadi pertimbangan penting dalam interpretasi

metrik evaluasi pada subbab berikutnya, karena akurasi keseluruhan berpotensi bias terhadap kelas mayoritas. Statistik deskriptif terhadap variabel numerik utama disajikan pada Tabel 4.

Tabel 4. Statistik Deskriptif Variabel Numerik Utama

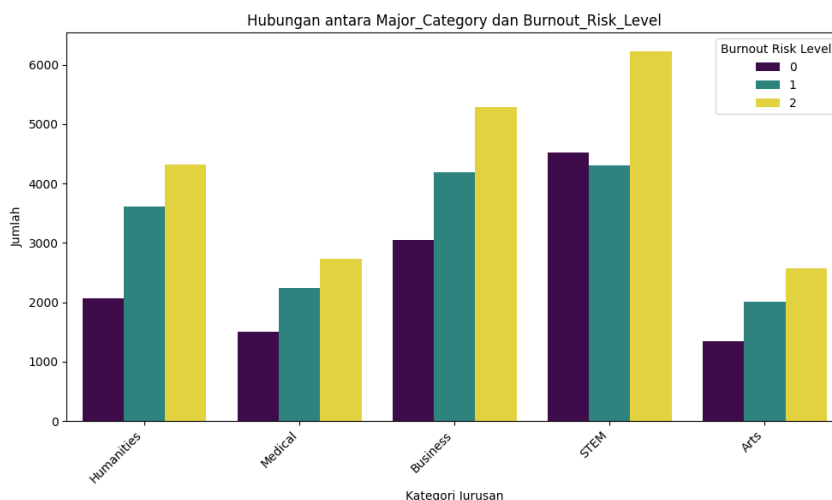
<i>Variabel</i>	<i>Mean</i>	<i>Std. Dev</i>	<i>Min</i>	<i>Maks</i>
<i>Pre_Semester_GPA</i>	3,15	0,48	1,18	4,00
<i>Weekly_GenAI_Hours</i>	8,43	8,27	0,00	40,00
<i>Traditional_Study_Hours</i>	11,21	5,16	1,00	35,86
<i>Perceived_AI_Dependency</i>	3,51	1,82	1,00	10,00
<i>Anxiety_Level_During_Exams</i>	4,27	2,14	1,00	10,00
<i>Post_Semester_GPA</i>	3,35	0,50	1,00	4,00
<i>Skill_Retention_Score</i>	75,80	13,28	10,78	100,00

Rerata penggunaan GenAI mingguan tercatat sebesar 8,43 jam dengan deviasi standar yang cukup besar (8,27), mengindikasikan variasi pola penggunaan yang lebar antar mahasiswa, mulai dari yang tidak menggunakan AI sama sekali hingga yang menggunakannya lebih dari 40 jam per minggu. Untuk melihat kecenderungan pola penggunaan GenAI pada masing-masing kelompok risiko burnout, dilakukan visualisasi distribusi menggunakan violin plot pada Gambar 2.



Gambar 2. Distribusi *Weekly_GenAI_Hours* berdasarkan *Burnout_Risk_Level*

Gambar 2 menunjukkan bahwa kelompok mahasiswa dengan risiko burnout High memiliki median penggunaan GenAI mingguan tertinggi (sekitar 13 jam/minggu), diikuti kelompok Medium (median sekitar 6 jam/minggu), dan kelompok Low dengan median penggunaan terendah (sekitar 3 jam/minggu). Pola ini secara visual mengindikasikan kecenderungan hubungan positif antara intensitas penggunaan Generative AI dengan tingkat risiko burnout mahasiswa. Analisis lanjutan terhadap hubungan kategori jurusan (*Major_Category*) dengan *Burnout_Risk_Level* divisualisasikan pada Gambar 3.



Gambar 3. Hubungan *Major_Category* dengan *Burnout_Risk_Level*

Secara umum, proporsi kelas Medium dan Low tampak konsisten lebih dominan dibandingkan kelas High pada hampir seluruh kategori jurusan. Jurusan STEM dan Business menunjukkan jumlah mahasiswa absolut tertinggi karena populasi sampel yang lebih besar, namun secara proporsional tidak ditemukan jurusan tertentu yang secara ekstrem mendominasi kelas High. Rerata Pre_Semester_GPA pada ketiga kelompok risiko juga relatif berdekatan, yaitu 3,08 (High), 3,20 (Low), dan 3,14 (Medium), yang mengindikasikan bahwa IPK sebelum semester berjalan bukan pembeda utama antar kelompok risiko burnout pada dataset ini.

3.2 Hasil Evaluasi Kinerja Model

Ketiga algoritma klasifikasi dilatih dan dievaluasi pada data uji (10.000 sampel), baik dalam kondisi baseline (parameter default) maupun setelah proses hyperparameter tuning menggunakan GridSearchCV. Ringkasan hasil evaluasi disajikan pada Tabel 5.

Tabel 5. Perbandingan Kinerja Model Sebelum dan Setelah Hyperparameter Tuning

Model	Akurasi	Presisi (W)	Recall (W)	F1-Score (W)
<i>Logistic Regression (Baseline)</i>	0,5383	0,5568	0,5383	0,5384
<i>Logistic Regression (Tuned)</i>	0,5381	0,5568	0,5381	0,5382
<i>Random Forest (Baseline)</i>	0,5238	0,5379	0,5238	0,5243
<i>Random Forest (Tuned)</i>	0,5327	0,5542	0,5327	0,5318
<i>XGBoost (Baseline)</i>	0,5196	0,5357	0,5196	0,5202
<i>XGBoost (Tuned)</i>	0,5362	0,5535	0,5362	0,5365

Hasil GridSearchCV menghasilkan kombinasi parameter terbaik untuk masing-masing model: Logistic Regression memperoleh parameter $C=0,1$, $penalty=l1$, $solver=saga$ dengan skor validasi silang 0,5367; Random Forest memperoleh $max_depth=10$, $min_samples_split=5$, $n_estimators=100$ dengan skor validasi silang 0,5320; dan XGBoost memperoleh $colsample_bytree=0,7$, $learning_rate=0,1$, $max_depth=3$, $n_estimators=200$, $subsample=0,7$ dengan skor validasi silang 0,5350.

Berdasarkan Tabel 5, terdapat tiga temuan utama. Pertama, model Logistic Regression baseline justru mencatatkan akurasi tertinggi secara keseluruhan (0,5383), sedikit lebih tinggi dibandingkan Logistic Regression setelah tuning (0,5381), Random Forest setelah tuning (0,5327), maupun XGBoost setelah tuning (0,5362). Hal ini mengindikasikan bahwa hubungan antara fitur prediktor dengan Burnout_Risk_Level pada dataset ini cenderung dapat ditangkap dengan baik oleh model linear sederhana, sehingga kompleksitas tambahan model non-linear tidak serta-merta memberikan keunggulan signifikan. Kedua, proses hyperparameter tuning memberikan peningkatan akurasi yang jelas pada Random Forest (dari 0,5238 menjadi 0,5327, naik 0,89 poin persentase) dan XGBoost (dari 0,5196 menjadi 0,5362, naik 1,66 poin persentase), namun justru menurunkan tipis akurasi Logistic Regression. Ketiga, selisih akurasi antar ketiga algoritma pada kondisi optimalnya masing-masing relatif kecil, berada pada rentang 0,5327 hingga 0,5383 (selisih sekitar 0,56 poin persentase), menunjukkan tidak ada satu algoritma yang secara dominan mengungguli algoritma lainnya pada permasalahan klasifikasi tiga kelas ini.

Hasil Classification Report untuk model Random Forest dan XGBoost setelah tuning sebagai dua model berbasis pohon dengan performa terbaik pascatuning disajikan pada Tabel 6 untuk melihat kinerja per kelas secara lebih rinci.

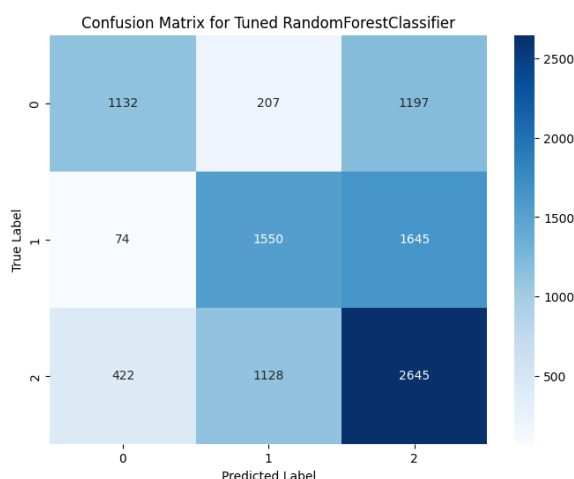
Tabel 6. Classification Report Model Random Forest dan XGBoost (Setelah Tuning)

Model	Kelas	Precision	Recall	F1-Score	Support
<i>Random Forest</i>	<i>High</i>	0,70	0,45	0,54	2.536
<i>Random Forest</i>	<i>Low</i>	0,54	0,47	0,50	3.269
<i>Random Forest</i>	<i>Medium</i>	0,48	0,63	0,55	4.195
<i>XGBoost</i>	<i>High</i>	0,69	0,47	0,56	2.536
<i>XGBoost</i>	<i>Low</i>	0,53	0,49	0,51	3.269
<i>XGBoost</i>	<i>Medium</i>	0,49	0,61	0,54	4.195

Tabel 6 menunjukkan pola yang konsisten pada kedua model: precision tertinggi terjadi pada kelas High (0,70 untuk Random Forest dan 0,69 untuk XGBoost), namun recall pada kelas tersebut paling rendah (0,45 dan 0,47). Sebaliknya, kelas Medium memiliki recall tertinggi (0,63 dan 0,61) tetapi precision terendah (0,48 dan 0,49). Pola ini mengindikasikan bahwa model cenderung lebih banyak memprediksi sampel sebagai kelas Medium, sehingga banyak sampel kelas High maupun Low yang sebenarnya turut terklasifikasi keliru (misclassified) ke dalam kelas Medium.

3.3 Analisis Confusion Matrix

Untuk memahami secara lebih spesifik letak kesalahan klasifikasi antar kelas, dilakukan visualisasi confusion matrix terhadap model Random Forest setelah tuning sebagai model dengan F1-Score makro tertinggi di antara model berbasis pohon. Hasil visualisasi disajikan pada Gambar 4.



Gambar 4. Confusion Matrix Model Random Forest (Setelah Tuning)

Nilai numerik pada matriks tersebut dirangkum kembali pada Tabel 7 agar lebih mudah diinterpretasikan.

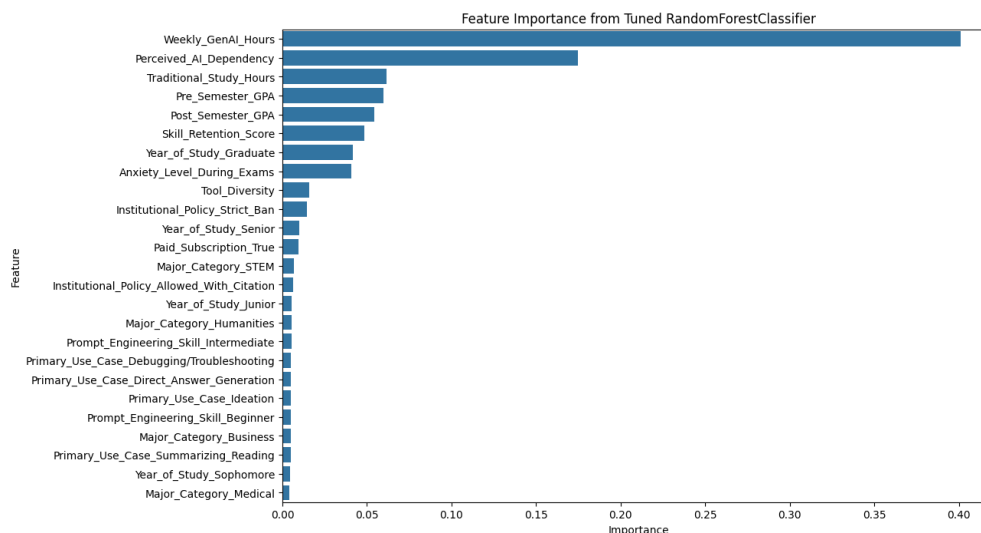
Tabel 7. Rincian Confusion Matrix Model Random Forest (Setelah Tuning)

Aktual \ Prediksi	High	Low	Medium
High	1.132	207	1.197
Low	74	1.550	1.645
Medium	422	1.128	2.645

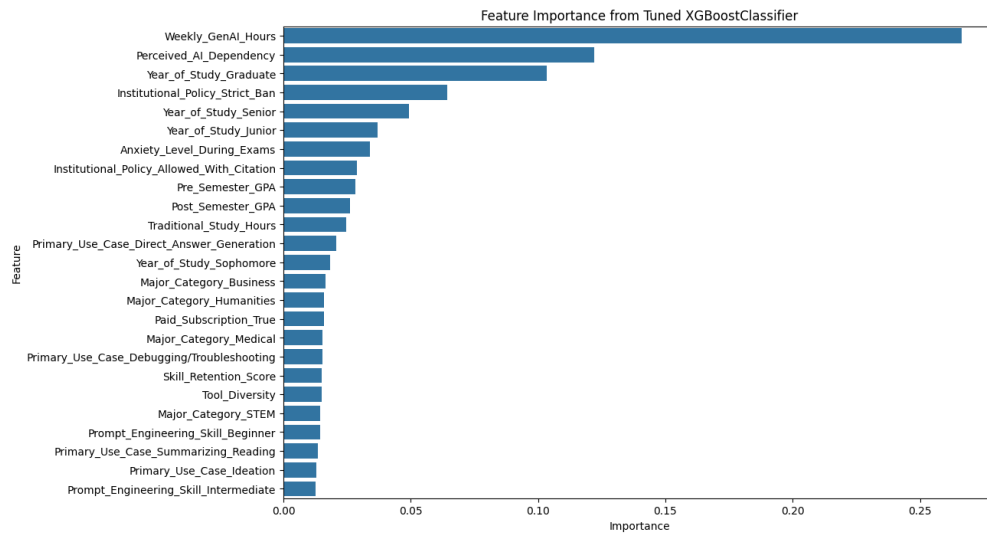
Dari Tabel 7 dapat diamati beberapa pola kesalahan klasifikasi yang menonjol. Pertama, dari 2.536 sampel kelas High yang sebenarnya, model hanya berhasil mengenali 1.132 sampel (44,6%) dengan benar, sedangkan 1.197 sampel (47,2%) diprediksi sebagai kelas Medium. Temuan ini penting karena kelas High merupakan kelompok mahasiswa dengan risiko burnout tertinggi yang paling krusial untuk dideteksi secara akurat dalam konteks intervensi dini. Kedua, kesalahan klasifikasi antara kelas Low dan High relatif kecil (hanya 74 dan 207 sampel pada masing-masing arah), menunjukkan model cukup baik membedakan kedua kelas pada ujung spektrum risiko yang berlawanan. Ketiga, kelas Medium menjadi tujuan misklasifikasi paling sering dari kedua kelas lainnya, sejalan dengan rendahnya precision kelas Medium yang telah dibahas pada Tabel 6. Pola ini mengindikasikan bahwa batas keputusan (decision boundary) antara kelas Medium dengan kelas High maupun Low pada dataset ini tidak terlalu tegas, kemungkinan karena distribusi fitur prediktor antar kelas yang saling beririsan (overlapping), sebagaimana juga tampak pada distribusi Weekly_GenAI_Hours pada Gambar 2.

3.4 Analisis Feature Importance

Bagian ini membedah variabel yang menjadi pemicu utama tingkat risiko burnout mahasiswa, diekstraksi melalui kapabilitas `feature_importances_` dari algoritma Random Forest dan XGBoost yang telah dioptimalkan. Hasil visualisasi feature importance untuk model Random Forest dan XGBoost masing-masing ditunjukkan pada Gambar 5 dan Gambar 6, sedangkan rekapitulasi sepuluh fitur teratas pada kedua model disajikan pada Gambar 7.



Gambar 5. Feature Importance Model Random Forest (Setelah Tuning)



Gambar 6. Feature Importance Model XGBoost (Setelah Tuning)

Peringkat	Random Forest	Importance	XGBoost	Importance
1	Weekly_GenAI_Hours	0,400	Weekly_GenAI_Hours	0,265
2	Perceived_AI_Dependency	0,175	Perceived_AI_Dependency	0,120
3	Traditional_Study_Hours	0,062	Year_of_Study_Graduate	0,103
4	Pre_Semester_GPA	0,059	Institutional_Policy_Strict_Ban	0,065
5	Post_Semester_GPA	0,054	Year_of_Study_Senior	0,049
6	Skill_Retention_Score	0,049	Year_of_Study_Junior	0,037
7	Year_of_Study_Graduate	0,041	Anxiety_Level_During_Exams	0,032
8	Anxiety_Level_During_Exams	0,040	Institutional_Policy_Allowed_With_Citation	0,028
9	Tool_Diversity	0,015	Pre_Semester_GPA	0,027
10	Institutional_Policy_Strict_Ban	0,013	Post_Semester_GPA	0,025

Gambar 7. Sepuluh Fitur dengan Nilai Importance Tertinggi

Temuan paling konsisten dari kedua model adalah bahwa Weekly_GenAI_Hours merupakan fitur dengan tingkat kepentingan tertinggi, baik pada Random Forest (importance 0,400) maupun XGBoost (importance 0,265), jauh meninggalkan fitur-fitur lainnya. Fitur Perceived_AI_Dependency secara konsisten berada pada peringkat kedua di kedua model. Temuan ini secara empiris mendukung hipotesis penelitian bahwa intensitas penggunaan dan ketergantungan terhadap Generative AI merupakan faktor utama yang memengaruhi tingkat risiko burnout mahasiswa. Perbedaan menarik terletak pada fitur-fitur peringkat menengah: Random Forest memberikan bobot yang relatif besar pada fitur akademik konvensional seperti Traditional_Study_Hours, Pre_Semester_GPA, dan Post_Semester_GPA, sedangkan XGBoost memberikan bobot lebih besar pada fitur kategorikal terkait tahun studi (Year_of_Study_Graduate, Year_of_Study_Senior, Year_of_Study_Junior) dan kebijakan institusi (Institutional_Policy_Strict_Ban). Perbedaan ini wajar terjadi karena kedua algoritma memiliki mekanisme pembentukan pohon keputusan dan fungsi split yang berbeda, namun keduanya tetap sepakat bahwa faktor terkait penggunaan AI merupakan dua faktor paling dominan.

3.5 Pengujian Hipotesis

Hipotesis H1(1) menetapkan bahwa hipotesis alternatif diterima apabila fitur Weekly_GenAI_Hours dan Perceived_AI_Dependency menempati posisi teratas pada ekstraksi feature importance. Berdasarkan Gambar 7, kedua fitur tersebut secara konsisten menempati peringkat pertama dan kedua pada model Random Forest maupun XGBoost, sehingga H1(1) dinyatakan diterima. Temuan ini mengonfirmasi bahwa durasi paparan GenAI dan tingkat ketergantungan psikologis merupakan dua determinan utama risiko burnout mahasiswa, jauh mendominasi faktor beban akademik konvensional.

Hipotesis H1(2) menyatakan bahwa hipotesis diterima jika terdapat selisih metrik evaluasi yang signifikan (>2%) antar ketiga algoritma pada kondisi yang sama. Berdasarkan Tabel 5, selisih akurasi terbesar pada kondisi baseline terjadi

antara Logistic Regression (0,5383) dan XGBoost (0,5196), yaitu sebesar 1,87 poin persentase, sementara pada kondisi setelah tuning selisih terbesar terjadi antara Logistic Regression (0,5381) dan Random Forest (0,5327), yaitu sebesar 0,54 poin persentase. Karena tidak satu pun selisih melampaui ambang batas 2%, maka H1(2) dinyatakan ditolak (H0(2) diterima), mengindikasikan bahwa ketiga algoritma memiliki kemampuan generalisasi yang relatif setara dalam memprediksi *Burnout_Risk_Level* pada dataset ini.

Hipotesis H1(3) menyatakan bahwa hipotesis diterima jika proses GridSearchCV menghasilkan peningkatan akurasi terukur (>1%) pada setidaknya satu model dibandingkan kondisi baseline. Berdasarkan Tabel 5, hyperparameter tuning berhasil meningkatkan akurasi XGBoost dari 0,5196 menjadi 0,5362, setara peningkatan 1,66 poin persentase yang melampaui ambang batas 1%, sehingga H1(3) dinyatakan diterima. Random Forest juga mengalami peningkatan signifikan sebesar 0,89 poin persentase, sementara Logistic Regression justru mengalami penurunan tipis 0,02 poin persentase, menunjukkan bahwa manfaat hyperparameter tuning lebih terasa pada model ensemble dibandingkan model linear yang ruang parameternya lebih sederhana.

Secara keseluruhan, meskipun akurasi prediksi ketiga model berada pada rentang moderat (52%-54%), pola hubungan antara intensitas penggunaan Generative AI dengan tingkat risiko burnout mahasiswa cukup konsisten teridentifikasi, baik melalui visualisasi EDA maupun analisis feature importance. Tingkat akurasi yang belum optimal kemungkinan dipengaruhi oleh distribusi fitur antar kelas yang saling beririsan, terutama tumpang tindih antara kelas Medium dengan kelas High dan Low, sebagaimana teridentifikasi pada analisis confusion matrix. Kondisi ini menjadi keterbatasan utama penelitian yang dapat menjadi landasan bagi penelitian selanjutnya untuk mengeksplorasi rekayasa fitur tambahan, teknik penyeimbangan kelas seperti SMOTE [12], maupun eksplorasi algoritma ensemble lain guna meningkatkan akurasi klasifikasi.

Temuan mengenai dominasi *Weekly_GenAI_Hours* dan *Perceived_AI_Dependency* sebagai prediktor utama burnout sejalan dengan hasil kajian sebelumnya yang menyoroti hubungan antara intensitas penggunaan teknologi digital dengan kelelahan belajar mahasiswa [5], maupun studi yang menunjukkan bahwa penggunaan alat analitik berbasis Machine Learning mampu mendeteksi kelompok mahasiswa berisiko stres akademik tinggi secara lebih objektif dibandingkan instrumen survei konvensional [6]. Dari sisi kebijakan, hasil ini memperkuat argumen bahwa ketergantungan psikologis terhadap ChatGPT dan alat GenAI sejenis, yang telah dilaporkan berasosiasi negatif dengan kesejahteraan akademis [9], perlu diimbangi dengan program literasi AI dan pendampingan konseling di tingkat institusi, bukan sekadar pembatasan akses teknologi semata. Dengan demikian, kontribusi utama penelitian ini terletak pada penyediaan bukti kuantitatif berskala besar (50.000 mahasiswa) yang menegaskan bahwa pola interaksi dengan Generative AI kini menjadi determinan burnout akademis yang setara penting, bahkan lebih dominan, dibandingkan faktor akademik tradisional yang selama ini menjadi fokus utama kebijakan institusi pendidikan.

4. KESIMPULAN

Berdasarkan hasil eksperimen pemodelan Machine Learning terhadap 50.000 data observasi mahasiswa dan evaluasi kinerja algoritma klasifikasi dalam memprediksi tingkat risiko burnout, dapat ditarik tiga kesimpulan utama. Pertama, pola interaksi mahasiswa dengan kecerdasan buatan terbukti menjadi faktor paling determinan dalam memicu kelelahan mental. Ekstraksi feature importance dari algoritma Random Forest dan XGBoost menunjukkan bahwa durasi penggunaan AI per minggu (*Weekly_GenAI_Hours*) dan tingkat ketergantungan psikologis (*Perceived_AI_Dependency*) secara konsisten menempati peringkat pertama dan kedua sebagai prediktor terkuat, jauh mendominasi faktor akademik tradisional seperti IPK dan jam belajar konvensional. Kedua, perbandingan performa ketiga model Machine Learning menunjukkan kemampuan generalisasi yang relatif setara; secara tak terduga, model linear sederhana Logistic Regression pada tahap baseline mencatatkan akurasi tertinggi (53,83%), sedikit mengungguli Random Forest (52,38%) dan XGBoost (51,96%), mengindikasikan bahwa data psikologis dalam penelitian ini tidak secara signifikan memperoleh keunggulan komputasi dari arsitektur model non-linear yang rumit. Ketiga, penerapan hyperparameter tuning dengan GridSearchCV berhasil menaikkan performa algoritma ensemble, dengan XGBoost meningkat menjadi 53,62% dan Random Forest menjadi 53,27%, namun akurasi prediksi maksimal secara keseluruhan tetap stagnan pada ambang batas sekitar 53,8% akibat tingginya varians dan sifat kelas target yang saling beririsan. Bagi institusi pendidikan, temuan ini menegaskan perlunya kebijakan pembatasan durasi penggunaan GenAI yang lebih terukur serta penyediaan layanan konseling untuk mengatasi ketergantungan AI. Bagi penelitian selanjutnya, disarankan penerapan teknik penyeimbangan kelas seperti SMOTE, rekayasa fitur tambahan yang mengombinasikan variabel penggunaan AI dengan metrik kecemasan, serta perluasan variabel psikologis dan demografis guna meningkatkan akurasi klasifikasi risiko burnout akademis di era kecerdasan buatan.

REFERENCES

- [1] E. Yeskuatov, L. K. Foo, and S.-L. Chua, "Detecting Burnout Among Undergraduate Computing Students with Supervised Machine Learning," *Healthcare*, vol. 13, no. 23, p. 3182, 2025, doi: 10.3390/healthcare13233182.
- [2] R. Pereira et al., "Burnout Risk Profiles in Psychology Students: An Exploratory Study with Machine Learning," *Behavioral Sciences*, vol. 15, no. 4, p. 505, 2025, doi: 10.3390/bs15040505.
- [3] G. Feher, K. Kapus, A. Tibold, Z. Banko, G. Berke, B. Gacs, I. Varadi, R. Nyulas, and A. Matuz, "Mental issues, internet addiction and quality of life predict burnout among Hungarian teachers: a machine learning analysis," *BMC Public Health*, vol. 24, p. 2322, 2024, doi: 10.1186/s12889-024-19797-9.
- [4] P. Sharma et al., "Hybrid Model for Early Detection of Student Burnout," *IEEE Transactions on Learning Technologies*, 2024.
- [5] N. A. Kuadey, C. Ankora, F. Tahiru, L. Bensah, C. C. M. Agbesi, and S. O. Bolatimi, "Using machine learning algorithms to examine the impact of technostress creators on student learning burnout and perceived academic performance," *International Journal of Information Technology*, vol. 16, no. 4, pp. 2467-2482, 2024, doi: 10.1007/s41870-023-01655-3.
- [6] H. Nellore, "Data Analytics and Machine Learning Approaches for Predicting Academic Stress and Enhancing Student Wellbeing," *International Journal of Advance Research and Innovation*, vol. 13, no. 4, pp. 23-29, 2025, doi: 10.69996/ijari.2025021.
- [7] R. P. Tapio, "Predicting Burnout in College Students: A Machine Learning Approach Using Decision Tree and Psychometric Data," *Asian Journal of Probability and Statistics*, vol. 27, no. 5, pp. 50-60, 2025, doi: 10.9734/ajpas/2025/v27i5754.
- [8] M. Jukiewicz, "How generative artificial intelligence transforms teaching and influences student wellbeing in future education," *Frontiers in Education*, vol. 10, 2025, doi: 10.3389/feduc.2025.1594572.
- [9] M. Besalti, "Harnessing Self-Control and AI: Understanding ChatGPT's Impact on Academic Wellbeing," *Behavioral Sciences*, vol. 15, no. 9, p. 1181, 2025, doi: 10.3390/bs15091181.
- [10] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, San Francisco, CA, USA, 2016, pp. 785-794, doi: 10.1145/2939672.2939785.
- [11] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001, doi: 10.1023/A:1010933404324.
- [12] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321-357, 2002, doi: 10.1613/jair.953.
- [13] J. Tian and R. Zhang, "Learners' AI dependence and critical thinking: The psychological mechanism of fatigue and the social buffering role of AI literacy," *Acta Psychologica*, vol. 260, p. 105725, 2025, doi: 10.1016/j.actpsy.2025.105725.
- [14] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825-2830, 2011.
- [15] D. M. W. Powers, "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation," *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37-63, 2011.