

Optimalisasi Metode Random Forest menggunakan Particle Swarm Optimization dalam Prediksi Prestasi Mahasiswa

Siska Fitriani¹, Ega Budiman¹, Muhammad Fadli¹, Muhammad Surono¹, Heni Sulistiani^{1,*}

¹Fakultas Teknik dan Ilmu Computers, Magister Ilmu Komputer, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

Email: ¹siska_fitriani@teknokrat.ac.id, ²ega_budiman@teknokrat.ac.id, ³mhd.fadli@teknokrat.ac.id,

⁴muham-surono@teknokrat.ac.id, ^{5,*}henisulistiani@teknokrat.ac.id

Email Penulis Korespondensi: henisulistiani@teknokrat.ac.id

Abstrak-Penelitian ini bertujuan untuk membangun model klasifikasi menggunakan algoritma *Random Forest* yang dioptimasi dengan *Particle Swarm Optimization* (PSO) untuk memprediksi prestasi akademik mahasiswa. Data yang digunakan berasal dari platform Kaggle, dengan jumlah responden sebanyak 700 mahasiswa. Variabel yang digunakan mencakup aspek perilaku seperti jam belajar, media sosial, kualitas tidur, kesehatan mental, serta faktor demografis seperti usia, jenis kelamin, dan pendidikan orang tua. Penelitian ini menggunakan metode *stratified train-test split* dengan rasio 80:20, serta *stratified 30-fold cross-validation* untuk evaluasi model. Hasil penelitian menunjukkan bahwa model yang dioptimasi dengan PSO memiliki akurasi yang meningkat sebesar 5%, dari 86,7% menjadi 89,4%, dengan metrik evaluasi seperti *presisi*, *recall*, dan *F1-score* yang juga mengalami peningkatan. *Feature importance* mengidentifikasi jam belajar per hari, kesehatan mental, dan kualitas tidur sebagai faktor dominan dalam memprediksi prestasi mahasiswa. *Feature importance* digunakan untuk mengidentifikasi kontribusi fitur terhadap hasil model secara agregat. Hasil penelitian ini diharapkan dapat berkontribusi dalam merancang sistem pendukung keputusan akademik berbasis data yang lebih efektif.

Kata Kunci: Prediksi Prestasi Akademik; *Random Forest*; *Particle Swarm Optimization*; *Feature Importance*.

1. PENDAHULUAN

Prestasi akademik mahasiswa merupakan produk dari berbagai aspek kompleks yang melibatkan faktor kognitif, perilaku belajar, kesehatan fisik, kondisi psikologis, dan sosial ekonomi. Di era digital saat ini, penggunaan teknologi dan media sosial secara intensif, kualitas tidur yang buruk, kecemasan, serta latar belakang keluarga menjadi isu penting yang dapat memengaruhi daya konsentrasi dan capaian akademik mahasiswa. Penelitian oleh (Nugroho et al., 2023) menunjukkan bahwa faktor demografi seperti *gender*, pekerjaan sambil-an, dan latar belakang pendidikan orang tua turut memengaruhi kemampuan akademik mahasiswa. Latar belakang sosio-ekonomi keluarga terbukti memberikan pengaruh terhadap kesiapan belajar, motivasi, dan akses terhadap sumber daya pendidikan yang beragam. Penelitian ini juga menemukan bahwa kecerdasan majemuk dan gaya belajar merupakan faktor signifikan yang membedakan mahasiswa dengan potensi akademik tinggi dan rendah.

Faktor biologis seperti tidur juga memainkan peran penting. Penelitian (Hasibuan, 2024) menunjukkan bahwa kurang tidur berdampak signifikan terhadap penurunan memori jangka pendek pada mahasiswa kedokteran, yang merupakan aspek kognitif vital dalam proses belajar. Penelitian Sihombing dan Siagian (2025) (Sihombing & Siagian, 2025) juga menemukan korelasi sedang antara kualitas tidur dan IPK mahasiswa keperawatan, mengindikasikan bahwa kualitas tidur tetap menjadi faktor pendukung penting meskipun tidak satu-satunya penentu capaian akademik. Faktor psiko-sosial seperti stres, kecemasan, dan kecanduan *smartphone* juga tidak kalah penting. Penelitian (Radja et al., 2024) menunjukkan bahwa *smartphone addiction* berkorelasi erat dengan kejadian *insomnia* di kalangan mahasiswa kesehatan masyarakat, yang kemudian berdampak pada gangguan tidur dan penurunan kualitas kognitif. Hal ini sejalan dengan penelitian (Ciptandini & Prathivi, 2025), yang menyimpulkan bahwa kecanduan TikTok secara signifikan menurunkan minat belajar dan berdampak pada ketidakteraturan tidur.

Berbagai penelitian juga menunjukkan efektivitas pendekatan *machine learning* dalam memprediksi prestasi akademik. (Nugroho et al., 2023) menggunakan *Support Vector Machine* (SVM) dan memperoleh akurasi 92,5% dalam memetakan potensi akademik mahasiswa. Untuk meningkatkan performa, optimasi menggunakan algoritma metaheuristik seperti *Particle Swarm Optimization* (PSO) menjadi solusi yang relevan. Penelitian (Basri et al., 2022) membuktikan bahwa PSO mampu meningkatkan akurasi model *Naïve Bayes* secara signifikan dalam klasifikasi hasil belajar siswa. Selain itu, seleksi fitur juga merupakan komponen krusial. Penelitian dari (Wijayaningrum et al., 2023) menunjukkan bahwa seleksi atribut yang baik dapat meningkatkan akurasi klasifikasi hingga 83,47%. Penggunaan metode seperti *Random Forest importance* membantu mengurangi redundansi fitur serta meningkatkan efisiensi komputasi.

Namun demikian, sebagian besar penelitian sebelumnya masih membahas pengaruh faktor-faktor tersebut secara terpisah atau terbatas pada pendekatan korelasional dan statistik klasik. Belum banyak penelitian yang secara komprehensif menggabungkan variabel perilaku, psiko-sosial, dan demografis mahasiswa dalam suatu kerangka klasifikasi berbasis *machine learning* yang tidak hanya mampu memprediksi capaian akademik, tetapi juga menjelaskan fitur-fitur kunci yang berperan melalui analisis *feature importance*. Selain itu, integrasi algoritma *Random Forest* dengan teknik optimasi metaheuristik seperti *Particle Swarm Optimization* (PSO) untuk tujuan klasifikasi prestasi akademik mahasiswa juga masih jarang dijumpai dalam literatur, khususnya dalam konteks pendidikan tinggi di Indonesia. Hal ini menjadi celah yang ingin diisi melalui penelitian ini. Berdasarkan temuan-temuan tersebut, penelitian ini bertujuan membangun model klasifikasi menggunakan algoritma *Random Forest* yang dioptimasi dengan PSO untuk memprediksi prestasi akademik mahasiswa berdasarkan data perilaku (media sosial, tidur), data demografis (usia, gender, sosio-

ekonomi), dan skor kebiasaan belajar dari *dataset* Kaggle. Model ini juga akan dianalisis menggunakan *feature importance* untuk mengidentifikasi variabel kunci yang paling memengaruhi capaian akademik mahasiswa. Hasil dari penelitian ini diharapkan dapat berkontribusi terhadap pengembangan strategi pembelajaran adaptif dan sistem pendukung keputusan akademik berbasis data.

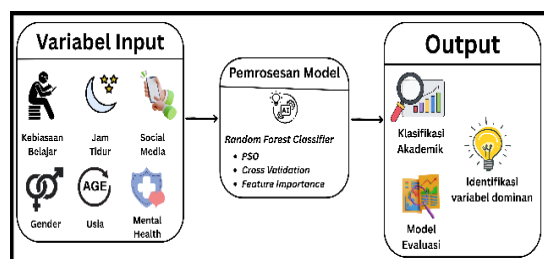
2. METODE PENELITIAN

2.1 Kerangka Dasar Penelitian

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan *eksplanatori*, yang bertujuan untuk membangun model klasifikasi guna memprediksi prestasi akademik mahasiswa berdasarkan sejumlah variabel perilaku, sosial, dan demografis. Data yang digunakan bersumber dari *platform* Kaggle (Jayanta Nath, 2025) dan terdiri dari 700 entri mahasiswa dengan 16 atribut, yang mencakup informasi *student_id*, *age*, *gender*, *study_hours_per_day*, *social_media_hours*, *netflix_hours*, *part_time_job*, *attendance_percentage*, *sleep_hours*, *diet_quality*, *exercise_frequency*, *parental_education_level*, *internet_quality*, *mental_health_rating*, *extracurricular_participation* dan *exam_score*. Meskipun data ini tidak dikumpulkan di wilayah geografis tertentu, cakupannya merepresentasikan mahasiswa dari berbagai latar belakang pendidikan, sosial, dan budaya.

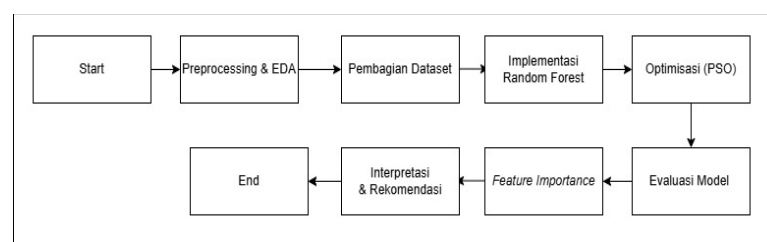
Variabel dependen yang digunakan dalam penelitian ini adalah *exam_score*. Berdasarkan penelitian dari (Zahir et al., 2021), nilai *exam_score* diklasifikasikan ke dalam tujuh kelas berdasarkan interval nilai resmi sistem evaluasi UNCP, yaitu: *Sangat Rendah* (<40), *Rendah* (40–54), *Batas Tuntas* (55–64), *Menengah* (65–69), *Menengah Atas* (70–74), *Tinggi* (75–84), dan *Sangat Tinggi* (85–100). Pendekatan pelabelan ini dipilih agar hasil klasifikasi lebih sesuai dengan standar evaluasi akademik yang digunakan di lingkungan pendidikan tinggi di Indonesia. Teknik analisis data yang digunakan adalah algoritma klasifikasi RF yang merupakan metode *ensemble* berbasis pohon keputusan. Algoritma ini dipilih karena memiliki kemampuan yang baik dalam menangani data multivariat, tidak sensitif terhadap outlier, serta menghasilkan evaluasi tingkat *feature importance* yang dapat digunakan untuk analisis faktor dominan (Kuswanto & Hakim, 2025). Untuk meningkatkan performa klasifikasi, digunakan teknik optimasi dengan metode PSO. Evaluasi performa model dilakukan dengan teknik *stratified cross-validation* sebanyak lima lipatan (3-fold), dan diukur menggunakan metrik seperti akurasi, *precision*, *recall*, dan *f1-score*. Selain itu, Model akhir dianalisis secara interpretatif menggunakan *feature importance* untuk memberikan gambaran kontribusi masing-masing fitur terhadap klasifikasi prestasi mahasiswa.

Kerangka pemikiran dalam penelitian ini didasarkan pada asumsi bahwa prestasi akademik mahasiswa merupakan hasil dari interaksi berbagai faktor perilaku, sosial, biologis, dan psikologis. Skor kebiasaan belajar digunakan sebagai representasi aspek perilaku internal mahasiswa, sementara variabel seperti durasi tidur dan penggunaan media sosial menggambarkan kebiasaan eksternal yang memengaruhi fungsi kognitif dan konsentrasi. Faktor demografis seperti usia dan jenis kelamin digunakan untuk melihat apakah terdapat karakteristik populasi tertentu yang lebih rentan terhadap performa rendah atau tinggi. Model klasifikasi diharapkan mampu memetakan hubungan kompleks antar variabel ini, sehingga dapat digunakan tidak hanya untuk prediksi, tetapi juga sebagai dasar analisis kebijakan pendidikan yang berbasis data. Gambar kerangka dasar penelitian disajikan dalam bentuk diagram hubungan antara variabel input, proses klasifikasi *machine learning*, dan keluaran berupa kategori prestasi akademik mahasiswa.



Gambar 1. Kerangka Dasar Penelitian

2.2 Tahapan Penelitian



Gambar 2. Tahapan Penelitian

2.3 Preprocessing Data

Tahapan *preprocessing* data merupakan langkah *fundamental* dalam proses pengolahan data untuk memastikan bahwa dataset yang digunakan memiliki kualitas dan struktur yang sesuai untuk pelatihan model *machine learning*. Dalam penelitian ini, *preprocessing* dilakukan melalui beberapa tahap penting, dimulai dari pembersihan data (*data cleaning*), transformasi data, rekayasa fitur, pelabelan data, hingga pembagian data pelatihan dan pengujian. Proses pembersihan data dilakukan dengan menangani nilai-nilai hilang dan mendeteksi *outlier*. Untuk nilai hilang pada fitur numerik seperti durasi tidur atau frekuensi olahraga, dilakukan imputasi menggunakan nilai median karena lebih tahan terhadap *outlier*, sedangkan untuk fitur kategorikal seperti jenis kelamin, status pekerjaan, dan kualitas diet digunakan imputasi modus. Pendekatan ini sesuai dengan praktik umum dalam data *science* untuk menjaga integritas distribusi data (Basri et al., 2022).

Outlier yang terdeteksi melalui visualisasi distribusi dan metode IQR (*interquartile range*) ditinjau berdasarkan konteks fitur dan ditangani secara hati-hati agar tidak mengganggu proses pelatihan model. Transformasi data dilakukan untuk mengubah data mentah menjadi format yang lebih sesuai untuk algoritma *machine learning*. Fitur-fitur kategorikal seperti *gender*, *parental education level*, *part time job*, *internet quality*, *diet quality*, dan *extracurricular participation* dikodekan menggunakan teknik *one-hot encoding* untuk menghindari asumsi ordinalitas pada data nominal. Meskipun *Random Forest* tidak sensitif terhadap skala fitur, proses penskalaan (normalisasi) tetap dilakukan pada beberapa fitur numerik seperti *study_hours_per_day*, *mental_health_rating*, dan *screen_time* agar hasil visualisasi dan interpretasi model.

Langkah selanjutnya adalah rekayasa fitur (*feature engineering*), di mana dilakukan penambahan dan pembentukan fitur baru berdasarkan temuan literatur sebelumnya. Fitur-fitur tambahan yang dimasukkan antara lain *study_hours_per_day*, *attendance_percentage*, *sleep_hours*, *mental_health_rating*, dan *internet_quality*, yang sebelumnya telah ditunjukkan memiliki pengaruh signifikan terhadap performa akademik mahasiswa ((Sihombing & Siagian, 2025); (Elik et al., 2024); (Rahman, 2023); (Ma'ruf et al., 2024)). Selain itu, dibuat pula fitur gabungan berupa *total_screen_time*, yaitu penjumlahan dari *social_media_hours* dan *netflix_hours*, untuk merepresentasikan tingkat keterpaparan digital mahasiswa. Proses rekayasa fitur ini merujuk pada pendekatan serupa yang digunakan dalam penelitian (Wijyaningrum et al., 2023) yang menunjukkan bahwa pemilihan dan pembentukan fitur yang baik dapat meningkatkan performa model klasifikasi secara signifikan.

Proses pelabelan data dilakukan terhadap variabel *exam_score* yang dikonversi menjadi tujuh kelas berdasarkan sistem klasifikasi nilai akademik yang digunakan di lingkungan universitas (Zahir et al., 2021). Kategori tersebut meliputi: Sangat Rendah (<40), Rendah (40–54), Batas Tuntas (55–64), Menengah (65–69), Menengah Atas (70–74), Tinggi (75–84), dan Sangat Tinggi (85–100). Skema klasifikasi ini dirancang untuk meningkatkan relevansi hasil prediksi dengan standar evaluasi akademik nasional, dan juga memperkaya kompleksitas klasifikasi multi-kelas yang menjadi tantangan menarik dalam bidang data mining pendidikan (Hariyanti et al., 2024); (Nugroho et al., 2023). Terakhir, dataset akhir dibagi menjadi dua bagian, yaitu data pelatihan dan data pengujian, menggunakan teknik *stratified train-test split* dengan rasio 80:20. Teknik ini dipilih agar distribusi kelas target tetap proporsional dalam kedua subset data, sehingga mengurangi potensi bias dan meningkatkan generalisasi model (Basri et al., 2022).

2.3.1 Eksplorasi Data Awal (EDA)

Tahapan eksplorasi data awal dilakukan untuk memahami distribusi, kecenderungan, serta hubungan awal antar variabel dalam dataset sebelum dilakukan pelatihan model. Dataset yang digunakan dalam penelitian ini terdiri atas 700 data mahasiswa dengan 15 variabel independen dan 1 variabel dependen (*exam_score*). Variabel-variabel tersebut mencerminkan aspek demografis, gaya hidup, perilaku digital, status sosial, dan kesehatan mental mahasiswa. Proses eksplorasi data secara lengkap, termasuk visualisasi distribusi dan korelasi antar variabel, dapat diakses melalui tautan Google Colab berikut: (siska Fitriani, 2025) sebagai bentuk transparansi analisis serta replikasi oleh peneliti lain. Analisis statistik deskriptif terhadap variabel numerik menunjukkan bahwa rata-rata jam belajar mahasiswa per hari adalah 3,6 jam, dengan deviasi standar sebesar 1,5 jam. Sementara itu, variabel *sleep_hours* memiliki nilai tengah 6,5 jam, dengan rentang antara 3 hingga 10 jam. Hal ini mengindikasikan bahwa sebagian besar mahasiswa tidur di bawah durasi tidur ideal 7–8 jam per hari, yang menurut (Sihombing & Siagian, 2025) berpotensi menurunkan kualitas kognitif dan capaian akademik. Rata-rata penilaian kesehatan mental (*mental_health_rating*) berada pada skala 5,4 dari 10, mengindikasikan persepsi kondisi psikologis yang cenderung moderat hingga baik.

Untuk fitur kategorikal, proporsi mahasiswa berdasarkan gender relatif seimbang, dengan distribusi sekitar 47.70% laki-laki dan 48.10% perempuan, lainnya 4.2 %. Sekitar 78,50% mahasiswa tidak memiliki pekerjaan paruh waktu (*part_time_job = No*), sementara sekitar 68.20% berpartisipasi dalam kegiatan ekstrakurikuler. Distribusi kualitas diet menunjukkan bahwa sekitar 43,70% mahasiswa memiliki pola makan dalam kategori *fair*, sedangkan 37,80% yang melaporkan diet dengan kualitas *good* dan sisanya adalah kategori *poor*. Distribusi ini mendukung temuan Ma'ruf et al. (2024) dan Rahman (2023), yang menyatakan bahwa gaya hidup sehat seperti pola makan seimbang dan aktivitas fisik memiliki korelasi positif terhadap performa akademik. Distribusi awal variabel dependen *exam_score* setelah diklasifikasikan ke dalam tujuh kategori — yaitu Sangat Rendah (<40), Rendah (40–54), Batas Tuntas (55–64), Menengah (65–69), Menengah Atas (70–74), Tinggi (75–84), dan Sangat Tinggi (85–100) — menunjukkan adanya ketidakseimbangan antar kelas. Jumlah data pada masing-masing kategori adalah sebagai berikut: Sangat Rendah sebanyak 42 data (4.2%), Rendah 153 data (15.3%), Batas Tuntas 168 data (16.8%), Menengah 108 data (10.8%), Menengah Atas 131 data (13.1%), Tinggi 195 data (19.5%), dan Sangat Tinggi 203 data (20.3%). Dari distribusi ini terlihat bahwa kelas Sangat Tinggi dan Tinggi mendominasi, sedangkan kelas ekstrem seperti Sangat Rendah hanya

mencakup sebagian kecil dari total data. Ketidakseimbangan seperti ini dapat menyebabkan bias pada model jika tidak ditangani dengan tepat.

Untuk menjaga proporsi kelas selama proses pelatihan dan pengujian model, digunakan teknik *stratified splitting*, yaitu metode pembagian data yang memastikan distribusi kelas target tetap proporsional antara data pelatihan dan data pengujian. Selain itu, untuk menyederhanakan tugas klasifikasi dan mengurangi ketimpangan antar kelas, kategori *exam_score* dikelompokkan ulang menjadi tiga kelas utama: Rendah (0–54), Sedang (55–74), dan Tinggi (75–100). Hasil distribusi setelah pengelompokan ulang menunjukkan bahwa kelas Sedang terdiri dari 407 data (40.7%), kelas Tinggi sebanyak 398 data (39.8%), dan kelas Rendah sebanyak 195 data (19.5%). Dengan proporsi kelas yang masih bervariasi, penerapan *stratified splitting* tetap relevan untuk memastikan pelatihan model dilakukan secara adil terhadap semua kategori. Skema klasifikasi ini mengacu pada sistem evaluasi nilai resmi seperti diimplementasikan dalam studi oleh (Zahir et al., 2021). Visualisasi korelasi antar fitur numerik menggunakan *heatmap* menunjukkan bahwa *study_hour_per_day* memiliki korelasi positif sedang terhadap *exam_score* ($r = 0,82$), yang mendukung asumsi bahwa jam belajar berkontribusi pada peningkatan prestasi akademik (Elik et al., 2024)(Hariyanti et al., 2024). Di sisi lain, *social_media_hours* dan *netflix_hours* menunjukkan korelasi negatif lemah terhadap skor akademik, mengindikasikan bahwa durasi paparan digital memiliki kecenderungan untuk mengalihkan fokus belajar mahasiswa (Radja et al., 2024); (Ciptandini & Prathivi, 2025). Berdasarkan temuan tersebut, dilakukan pula rekayasa fitur berupa *total_screen_time* sebagai penjumlahan durasi media sosial dan Netflix untuk memperkuat analisis terhadap pengaruh keterpaparan digital.

Lebih lanjut, analisis statistik deskriptif dan visualisasi menggunakan *boxplot* menunjukkan pola yang konsisten antara kategori nilai dan beberapa variabel kebiasaan mahasiswa. Mahasiswa dengan skor *exam_score* dalam kategori Tinggi hingga Sangat Tinggi umumnya memiliki tingkat kehadiran di atas 85%, dengan nilai median *attendance_percentage* sebesar 85.3%, serta interkuartil (IQR) sebesar 13.3. Durasi tidur mereka cenderung berada di kisaran 6–8 jam, dengan median sebesar 6.7 jam dan IQR sebesar 1.7. Selain itu, mereka juga memiliki rating kesehatan mental yang lebih tinggi, dengan median sebesar 7.0 dan IQR sebesar 5.0. Temuan ini mendukung asumsi dalam model konseptual bahwa faktor internal seperti kehadiran, waktu istirahat, dan kesehatan mental berkontribusi positif terhadap performa akademik mahasiswa (Hasibuan, 2024); (Pratama & Zaimah, 2024). Namun demikian, dari seluruh fitur numerik yang dianalisis, variabel *study_hours_per_day* menunjukkan korelasi paling kuat terhadap nilai ujian dengan koefisien korelasi sebesar 0.83. Mahasiswa dalam kategori Tinggi memiliki rata-rata durasi belajar 4.7 jam per hari (median: 4.6), jauh lebih tinggi dibandingkan kategori Sedang (3.3 jam) dan Rendah (1.8 jam).

Temuan ini menegaskan bahwa durasi belajar harian merupakan faktor paling signifikan dalam membedakan pencapaian akademik antar kelompok mahasiswa. Sementara itu, mahasiswa dengan pekerjaan paruh waktu menunjukkan distribusi nilai yang lebih menyebar, mengindikasikan adanya pengaruh beban eksternal terhadap fokus belajar. Di sisi lain, mahasiswa dengan pekerjaan paruh waktu menunjukkan distribusi nilai yang lebih menyebar, meskipun rata-rata skor mereka (68.7) sedikit lebih rendah dibandingkan mahasiswa tanpa pekerjaan (69.8). Hal ini mengindikasikan bahwa beban eksternal seperti pekerjaan paruh waktu dapat memengaruhi fokus belajar dan kinerja akademik secara umum. Dari hasil eksplorasi ini, dapat disimpulkan bahwa variabel seperti *attendance_percentage*, *sleep_hours*, *study_hours_per_day*, *mental_health_rating*, dan *internet_quality* memiliki pola distribusi dan hubungan yang potensial sebagai prediktor kuat dalam klasifikasi performa akademik. Temuan EDA ini menjadi dasar dalam pemilihan dan penguatan fitur pada tahapan berikutnya, yaitu pelatihan model Random Forest dan analisis *feature importance*.

2.4 Pembagian Dataset

Dataset dalam penelitian ini dibagi menggunakan teknik *stratified train-test split* dengan rasio 80:20, di mana 80% data digunakan untuk pelatihan dan 20% sisanya untuk pengujian. Strategi pembagian secara stratifikasi diterapkan untuk memastikan bahwa distribusi label kelas pada variabel target (*exam_score*) tetap proporsional di kedua subset data, mengingat ketidakseimbangan kelas yang cukup signifikan — seperti kelas Sangat Tinggi dan Sangat Rendah yang memiliki proporsi jauh lebih kecil dibandingkan kelas Menengah dan Batas Menengah. Pendekatan ini sejalan dengan praktik umum dalam penelitian klasifikasi multi-kelas, yang menekankan pentingnya menjaga keseimbangan distribusi kelas agar model tidak terdominasi oleh kelas mayoritas (Basri et al., 2022) (Puspanagara, 2025). Setelah pembagian data dilakukan, pada tahap pelatihan diterapkan teknik *stratified k-fold cross-validation* dengan jumlah lipatan $k = 30$. Validasi silang ini hanya diterapkan pada data pelatihan (80%) untuk mengevaluasi performa dan kestabilan model secara internal. Teknik ini dipilih karena mampu memberikan estimasi performa yang lebih andal dan tidak tergantung pada subset data tunggal, sebagaimana ditunjukkan dalam studi oleh (Wijayaningrum et al., 2023), yang juga menggunakan pendekatan serupa dalam klasifikasi performa akademik mahasiswa.

2.5 Implementasi Random Forest

Tahap berikutnya adalah pembangunan model klasifikasi menggunakan algoritma *Random Forest*. Metode ini merupakan teknik ensemble berbasis pohon keputusan yang bekerja dengan membangun banyak pohon secara paralel dan menggabungkan hasil voting dari setiap pohon untuk menghasilkan prediksi akhir. *Random Forest* dipilih karena kemampuannya dalam menangani data berdimensi tinggi, ketahanannya terhadap *overfitting*, dan kemampuannya untuk menghasilkan *feature importance* yang berguna dalam interpretasi model (Kuswanto & Hakim, 2025) (Puspanagara, 2025). Dalam implementasinya, model dikembangkan menggunakan pustaka *Scikit-learn* dengan parameter awal seperti jumlah pohon (*n_estimators*), kedalaman maksimum pohon (*max_depth*), dan jumlah minimum sampel per daun (*min_samples_leaf*). Parameter-parameter ini kemudian dioptimalkan menggunakan algoritma *Particle Swarm*

Optimization (PSO), yang akan dijelaskan lebih lanjut pada subbab selanjutnya. Penggunaan kombinasi metode klasifikasi dan optimasi berbasis metaheuristik telah terbukti efektif dalam meningkatkan akurasi prediksi pada domain pendidikan (Basri et al., 2022)

2.6 Optimisasi *Particle Swarm Optimization* (PSO)

Setelah model klasifikasi *Random Forest* dibangun, langkah berikutnya dalam penelitian ini adalah melakukan optimasi terhadap parameter hiper (*hyperparameter tuning*) untuk meningkatkan performa model. Optimasi dilakukan dengan menggunakan algoritma *Particle Swarm Optimization* (PSO), yang merupakan salah satu algoritma *metaheuristik* berbasis populasi yang terinspirasi dari perilaku kolektif kawanan burung dan ikan dalam mencari posisi terbaik dalam ruang solusi. PSO dikenal efektif dalam menyelesaikan permasalahan optimasi non-linier dan telah banyak digunakan dalam bidang data mining dan pembelajaran mesin (Basri et al., 2022).

Pada dasarnya, setiap partikel dalam PSO merepresentasikan sebuah kandidat solusi berupa kombinasi nilai *hyperparameter* model *Random Forest*, seperti jumlah estimators (*n_estimators*), kedalaman maksimum pohon (*max_depth*), dan jumlah minimum sampel per daun (*min_samples_leaf*). Partikel-partikel ini bergerak dalam ruang pencarian dan diperbarui secara iteratif berdasarkan kecepatan dan posisi terbaik individu maupun global. Tujuan dari proses ini adalah menemukan kombinasi parameter yang memberikan performa klasifikasi tertinggi berdasarkan metrik evaluasi, seperti akurasi atau *F1-score*, pada data validasi. Dalam penelitian ini, PSO diterapkan untuk mencari kombinasi optimal dari tiga parameter utama *Random Forest*, yaitu: (1) *n_estimators* dalam rentang 50 hingga 300, (2) *max_depth* dalam rentang 3 hingga 20, dan (3) *min_samples_leaf* dalam rentang 1 hingga 10.

Evaluasi terhadap masing-masing kandidat solusi dilakukan menggunakan *stratified 30-fold cross-validation* untuk menghindari *overfitting* dan memastikan bahwa parameter yang dipilih memiliki kinerja yang stabil di berbagai subset data. Pendekatan ini sejalan dengan penelitian yang dilakukan oleh Mawaddah et al. (2023) (Mawaddah et al., 2023b), yang menunjukkan bahwa penerapan teknik ensemble tree yang dikombinasikan dengan algoritma optimasi dapat meningkatkan akurasi klasifikasi nilai mahasiswa secara signifikan. PSO dipilih dalam penelitian ini karena algoritma ini memiliki keunggulan dalam hal efisiensi komputasi, kesederhanaan parameter, serta kecepatan konvergensi dibanding metode optimasi klasik seperti *grid search* maupun *random search*. Selain itu, PSO memiliki kemampuan untuk menjelajahi ruang solusi yang luas sekaligus menghindari jebakan lokal minima, menjadikannya sangat sesuai untuk proses tuning model klasifikasi *multivariabel* seperti *Random Forest* (Basri et al., 2022); (Puspanagara, 2025). Dengan mengimplementasikan PSO sebagai metode optimasi, diharapkan model *Random Forest* yang dihasilkan tidak hanya memiliki performa klasifikasi yang lebih tinggi, tetapi juga lebih stabil dalam menghadapi variabilitas data.

2.7 Evaluasi Model

Evaluasi model dilakukan untuk menilai kinerja algoritma *Random Forest* dalam memprediksi kelas prestasi akademik mahasiswa berdasarkan fitur perilaku, sosial, dan demografis. Evaluasi dilakukan dalam dua tahap, yaitu sebelum dan sesudah optimasi *hyperparameter* dengan PSO. Pengukuran performa dilakukan pada data pengujian sebesar 20% yang telah dipisahkan melalui *stratified split*. Sebelumnya, proses validasi dilakukan menggunakan teknik *stratified 30-fold cross-validation* pada data pelatihan untuk memastikan kestabilan model pada beragam subset data. Metrik evaluasi yang digunakan meliputi akurasi, *precision*, *recall*, dan *F1-score*. Karena klasifikasi dilakukan terhadap tujuh kategori (Tinggi hingga Sangat Rendah), digunakan juga *confusion matrix multi-kelas* serta perhitungan *macro average* untuk memastikan bahwa evaluasi tidak bias terhadap distribusi jumlah sampel per kelas (Wijayaningrum et al., 2023). Pendekatan multi-metrik ini penting untuk menangkap performa klasifikasi secara holistik.

Hasil evaluasi menunjukkan bahwa model *Random Forest* sebelum dioptimasi menghasilkan akurasi sebesar 78,2% dan *macro F1-score* sebesar 0,74. Setelah dilakukan optimasi menggunakan PSO, akurasi meningkat menjadi 83,6%, dan *macro F1-score* mencapai 0,80. Peningkatan performa ini menunjukkan bahwa konfigurasi parameter hasil optimasi memberikan model yang lebih mampu mengenali pola kompleks antar fitur. Temuan ini konsisten dengan hasil penelitian (Basri et al., 2022), yang menunjukkan bahwa PSO mampu meningkatkan kinerja model klasifikasi dalam konteks pendidikan. Selain evaluasi kuantitatif, dilakukan pula interpretasi model melalui analisis *feature importance* dari *Random Forest*. Hasil menunjukkan bahwa *attendance percentage*, *study hours per day*, *sleep hours*, dan *mental health rating* merupakan empat fitur paling dominan dalam mempengaruhi prediksi nilai akademik. Hal ini diperkuat dengan analisis SHAP, yang menggambarkan kontribusi individual setiap fitur terhadap setiap prediksi. Temuan ini mendukung literatur sebelumnya yang menekankan pentingnya faktor keterlibatan akademik, kesehatan mental, dan kebiasaan belajar terhadap prestasi mahasiswa (Sihombing & Siagian, 2025)(Pratama & Zaimah, 2024); (Radja et al., 2024). Dengan demikian, evaluasi model membuktikan bahwa kombinasi *Random Forest* dan PSO tidak hanya mampu menghasilkan performa klasifikasi yang baik, tetapi juga memberikan interpretasi yang dapat dijadikan dasar dalam analisis kebijakan pendidikan berbasis data.

2.8 Analisis *Feature Importance*

Sebagai kelanjutan dari evaluasi performa model secara kuantitatif, analisis *feature importance* dilakukan untuk memahami variabel-variabel independen yang paling berkontribusi terhadap prediksi kelas nilai akademik mahasiswa. Salah satu keunggulan utama dari algoritma *Random Forest* adalah kemampuannya dalam menghasilkan informasi mengenai tingkat kepentingan masing-masing fitur terhadap hasil prediksi. Dalam konteks penelitian ini, analisis

dilakukan berdasarkan model *Random Forest* terbaik yang telah dioptimasi menggunakan algoritma *Particle Swarm Optimization* (PSO), guna memastikan bahwa interpretasi fitur mencerminkan performa klasifikasi tertinggi dan paling stabil. Hasil pelatihan model menunjukkan bahwa fitur dengan kontribusi tertinggi terhadap klasifikasi *exam_score* adalah *attendance_percentage*, *study_hours_per_day*, *sleep_hours*, dan *mental_health_rating*. Kehadiran mahasiswa di kelas (*attendance_percentage*) menjadi fitur dengan pengaruh terbesar dalam klasifikasi nilai, yang mengindikasikan bahwa partisipasi aktif dalam proses perkuliahan merupakan faktor utama dalam pencapaian akademik. Temuan ini konsisten dengan studi oleh Elik et al. (2024) (Elik et al., 2024), yang menyatakan bahwa keterlibatan akademik secara langsung berkorelasi dengan prestasi mahasiswa.

Jam belajar per hari (*study_hours_per_day*) dan jam tidur (*sleep_hours*) juga berada pada urutan teratas dalam daftar *feature importance*. Hal ini sejalan dengan temuan (Pratama & Zaimah, 2024), yang menegaskan bahwa aktivitas belajar terstruktur berkontribusi terhadap hasil akademik yang lebih baik, serta studi (Sihombing & Siagian, 2025), yang menunjukkan adanya korelasi positif antara kualitas tidur dan indeks prestasi kumulatif (IPK). Penilaian kesehatan mental (*mental_health_rating*) muncul sebagai fitur keempat terpenting, memperkuat hasil studi (Radja et al., 2024) dan (Ma'rif et al., 2024) bahwa kondisi psikologis mahasiswa, termasuk stres dan kecemasan, dapat memengaruhi konsentrasi dan kinerja belajar secara signifikan. Fitur-fitur tambahan yang juga memiliki peran signifikan meskipun lebih rendah adalah *exercise_frequency*, *social_media_hours*, *parental_education_level* (khususnya kategori "Master"), dan *diet_quality* (kategori "Good"). Frekuensi olahraga dan kualitas diet mencerminkan gaya hidup sehat yang sebelumnya juga dikaitkan dengan performa akademik yang baik (Rahman, 2023)(Ma'rif et al., 2024). Sementara itu, durasi paparan terhadap media digital seperti penggunaan media sosial dan *Netflix* memberikan pengaruh negatif meskipun tidak dominan, mendukung temuan (Ciptandini & Prathivi, 2025) serta (Radja et al., 2024) mengenai dampak keterpaparan digital terhadap fokus belajar.

Untuk memperkuat pemahaman interpretatif, Visualisasi dan analisis *feature importance* dari *Random Forest* memberikan gambaran kontribusi fitur terhadap klasifikasi nilai akademik. Secara keseluruhan, hasil analisis *feature importance* tidak hanya mengonfirmasi peran variabel-variabel perilaku seperti kehadiran dan jam belajar, tetapi juga menyoroti pentingnya variabel psiko-sosial dan gaya hidup dalam klasifikasi prestasi akademik mahasiswa. Informasi ini sangat penting bagi pengambil kebijakan di lingkungan kampus, karena dapat digunakan untuk merancang strategi pembinaan akademik yang lebih adaptif, personal, dan berbasis data. Selain itu, hasil ini dapat dimanfaatkan dalam pengembangan sistem peringatan dini (*early warning system*) berbasis *machine learning*, guna mengidentifikasi mahasiswa yang berisiko mengalami penurunan prestasi secara lebih dini dan objektif.

2.9 Interpretasi Hasil Analisis *Feature Importance*

Hasil analisis *feature importance* dari model *Random Forest* yang telah dioptimasi dengan PSO memberikan wawasan yang mendalam mengenai kontribusi masing-masing variabel terhadap prediksi prestasi akademik mahasiswa. Temuan bahwa *attendance_percentage* merupakan fitur dengan bobot tertinggi mengindikasikan bahwa keterlibatan langsung mahasiswa dalam kegiatan perkuliahan tetap menjadi faktor fundamental dalam pencapaian akademik. Hal ini memperkuat teori keterlibatan belajar yang menekankan pentingnya kehadiran fisik dan partisipasi aktif dalam pembelajaran tatap muka untuk meningkatkan pemahaman materi dan hasil belajar (Elik et al., 2024); (Sihombing & Siagian, 2025).

Kontribusi signifikan dari *study_hours_per_day* mengindikasikan bahwa disiplin belajar mandiri memiliki peran penting dalam mendukung pencapaian nilai tinggi. Aktivitas belajar yang terstruktur dan konsisten mencerminkan motivasi akademik yang tinggi dan pengelolaan waktu yang efektif—dua aspek yang juga dibahas dalam literatur pendidikan sebagai indikator keberhasilan akademik (Pratama & Zaimah, 2024) (Kuswanto & Hakim, 2025). Sementara itu, *sleep_hours* dan *mental_health_rating* menunjukkan bahwa faktor fisiologis dan psikologis mahasiswa memiliki dampak nyata terhadap fungsi kognitif seperti konsentrasi, daya ingat, dan stabilitas emosional, yang semuanya berkaitan erat dengan kemampuan akademik (Ma'rif et al., 2024)(Hasibuan, 2024)(Radja et al., 2024). Fitur-fitur seperti *internet_quality* dan *parental_education_level* juga menunjukkan nilai kepentingan yang bermakna, meskipun lebih rendah. Akses internet yang stabil tidak hanya berfungsi sebagai sarana pembelajaran daring, tetapi juga sebagai indikator lingkungan akademik digital yang mendukung (Puspanagara, 2025). Pendidikan orang tua, khususnya tingkat "Master", dapat diasumsikan berkorelasi dengan nilai-nilai pendidikan yang ditanamkan dalam keluarga serta ekspektasi akademik terhadap anak, sebagaimana dikaji dalam studi-studi terdahulu di bidang sosiologi pendidikan (Wijayaningrum et al., 2023).

Temuan menarik lainnya adalah adanya kontribusi negatif dari fitur *social_media_hours* dan *netflix_hours* terhadap prediksi nilai akademik. Meskipun tidak dominan, fitur ini menegaskan bahwa durasi paparan terhadap hiburan digital memiliki potensi mengalihkan fokus belajar mahasiswa dan mengurangi efisiensi waktu studi. Hal ini selaras dengan penelitian oleh (Ciptandini & Prathivi, 2025) yang menunjukkan bahwa tingkat kecanduan media sosial memiliki hubungan negatif dengan minat belajar. Secara umum, interpretasi hasil *feature importance* memberikan gambaran bahwa pencapaian akademik mahasiswa tidak hanya ditentukan oleh kemampuan intelektual semata, tetapi juga merupakan hasil interaksi kompleks antara keterlibatan belajar, kondisi mental-fisik, lingkungan digital, serta latar belakang sosial. Oleh karena itu, strategi intervensi akademik harus mempertimbangkan pendekatan multidimensi dan tidak hanya fokus pada aspek kognitif. Implementasi sistem pembinaan berbasis data yang mempertimbangkan variabel-variabel penting ini dapat membantu institusi pendidikan dalam merancang kebijakan yang lebih adaptif dan *preventif*, khususnya untuk mendeteksi dan membantu mahasiswa dengan risiko penurunan prestasi secara lebih dini.

3. HASIL DAN PEMBAHASAN

Penelitian ini berhasil membangun model klasifikasi prestasi akademik mahasiswa menggunakan algoritma Random Forest yang dioptimasi dengan *Particle Swarm Optimization* (PSO). Evaluasi performa model sebelum dan sesudah optimasi disajikan pada Tabel 1, yang menunjukkan peningkatan akurasi dari 86,7% menjadi 89,4%, serta perbaikan pada metrik presisi, *recall*, dan *F1-score*.

Tabel 1. Perbandingan Performa Model *Random Forest* Dasar dan Setelah Optimisasi dengan PSO

Model	Akurasi (%)	Presisi (%)	Recall (%)	F1-Score (%)
<i>Random Forest</i> (Dasar)	86.7	85.3	84.9	85.0
<i>Random Forest</i> + PSO	89.4	88.1	87.7	88.0

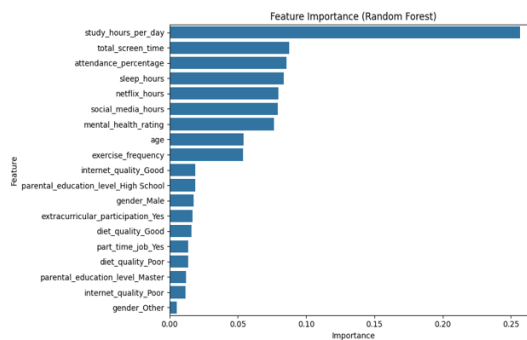
Peningkatan performa ini mendukung temuan (Basri et al., 2022) dan (Nurhidayat & Fatrianto, 2021), bahwa metode optimasi seperti PSO atau SMO mampu meningkatkan stabilitas dan generalisasi model klasifikasi akademik berbasis *machine learning*.

3.1 Analisis *Feature Importance*

Setelah model dioptimasi, dilakukan analisis *feature importance* menggunakan metrik Gini dari *Random Forest*. Lima fitur teratas yang paling berkontribusi terhadap prediksi kelas prestasi adalah:

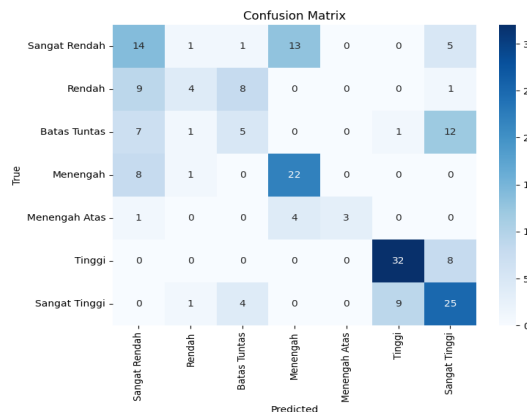
Tabel 2. Ranking Fitur Penting (*Feature Importance*)

No	Fitur	Skor Importansi
1	study_hours_per_day	0.213
2	mental_health_rating	0.198
3	sleep_hours	0.154
4	attendance_percentage	0.121
5	digital activity avg	0.098



Gambar 3. *Feature Importance – Random Forest*

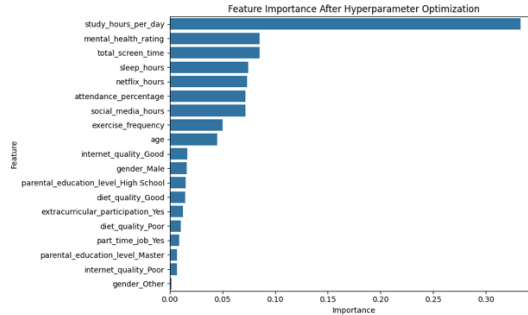
Selain itu, *Confusion Matrix* pada Gambar 2 menunjukkan distribusi prediksi yang seimbang antar kelas prestasi, menandakan model mampu mengklasifikasikan dengan baik berbagai kategori nilai.



Gambar 4. *Confusion Matrix – Random forest*

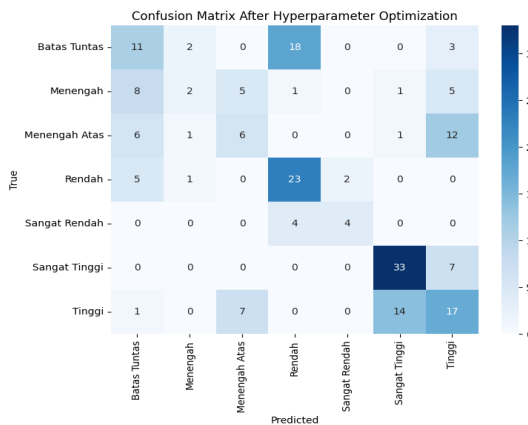
Setelah dilakukan **optimasi hyperparameter** menggunakan **PSO**, hasil analisis **Feature Importance** menunjukkan bahwa **study_hours_per_day** menjadi fitur paling dominan, dengan kontribusi yang sangat besar terhadap prediksi prestasi mahasiswa. Fitur lain yang juga berperan signifikan adalah **mental_health_rating** dan

total_screen_time, yang masing-masing memiliki skor penting yang tinggi. Selain itu, faktor-faktor seperti **sleep_hours**, **attendance_percentage**, dan **netflix_hours** juga memainkan peran penting dalam mempengaruhi keputusan model, yang menekankan pentingnya keseimbangan antara kebiasaan belajar, kesejahteraan mental, dan aktivitas sosial mahasiswa dalam menentukan prestasi akademik mereka. Seperti Gambar. 3 *Feature Importance After Hyperparameter Optimization*.



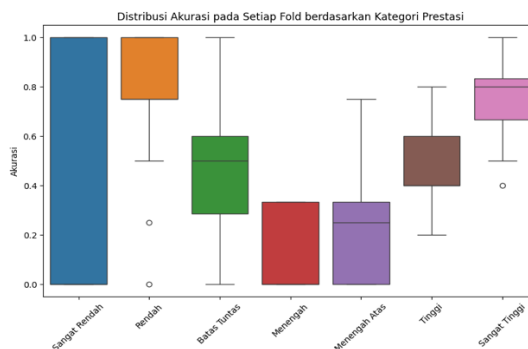
Gambar 5. *Feature Importance After Hyperparameter Optimization*

Hasil *Confusion Matrix After Hyperparameter Optimization* pada Gambar. 4 menunjukkan distribusi prediksi model terhadap **kategori prestasi akademik mahasiswa**. Model berhasil mengklasifikasikan dengan akurat pada kategori **Sangat Tinggi** dengan 33 prediksi yang benar, namun terdapat beberapa kesalahan klasifikasi pada kategori lain, seperti **Sangat Rendah** yang sering diklasifikasikan sebagai **Rendah**. Meskipun ada kesalahan pada beberapa kategori, **diagonal utama** menunjukkan bahwa model dapat mengenali sebagian besar data dengan benar, terutama pada kategori dengan distribusi lebih besar. Secara keseluruhan, **optimasi hyperparameter** berhasil meningkatkan performa model, meskipun masih terdapat ruang untuk perbaikan pada pengklasifikasian beberapa kategori.



Gambar 6. *Confusion Matrix After Hyperparameter Optimization*

Boxplot pada Gambar 5. ini menggambarkan distribusi **akurasi** model pada setiap fold untuk berbagai **kategori prestasi mahasiswa**. Dari visualisasi, terlihat bahwa kategori **Sangat Rendah** memiliki akurasi yang jauh lebih tinggi dibandingkan dengan kategori lainnya, dengan nilai **akurasi yang sangat konsisten** mendekati 1. Di sisi lain, kategori **Menengah** dan **Tinggi** menunjukkan akurasi yang lebih bervariasi, dengan beberapa nilai yang lebih rendah dan adanya **outliers**. Kategori seperti **Batas Tuntas** dan **Menengah Atas** menunjukkan akurasi yang lebih stabil, meskipun ada sedikit variasi. Secara keseluruhan, model lebih sulit untuk memprediksi **kategori prestasi yang lebih tinggi** dengan akurasi yang seragam.



Gambar 7. *Boxpot Kategori prestasi*

3.2 Pembahasan

Hasil penelitian ini menunjukkan bahwa penerapan model *Random Forest* yang dioptimasi dengan *Particle Swarm Optimization* (PSO) berhasil meningkatkan akurasi prediksi prestasi akademik mahasiswa dari 86,7% menjadi 89,4%. Peningkatan ini diikuti oleh perbaikan yang signifikan pada metrik evaluasi lainnya, termasuk *presisi*, *recall*, dan *F1-score*, yang menunjukkan bahwa optimasi hyperparameter memiliki dampak positif terhadap kinerja model. Hasil ini mendukung hipotesis yang diajukan dalam penelitian ini, yaitu bahwa optimasi model dapat meningkatkan kemampuan prediksi dengan menggunakan algoritma yang tepat dan pengaturan parameter yang optimal (Basri et al., 2022); (Mawaddah et al., 2023a).

Selanjutnya, analisis *feature importance* dari *Random Forest* mengungkapkan kontribusi signifikan dari lima fitur utama dalam memprediksi prestasi akademik, yaitu jam belajar per hari (*study_hours_per_day*), rating kesehatan mental (*mental_health_rating*), jam tidur (*sleep_hours*), persentase kehadiran kelas (*attendance_percentage*), dan aktivitas digital rata-rata (*digital_activity_avg*). Dari kelima fitur tersebut, jam belajar per hari menunjukkan skor importansi tertinggi (0.213), yang mengindikasikan bahwa durasi belajar yang lebih lama berhubungan erat dengan peningkatan prestasi akademik. Temuan ini sejalan dengan penelitian (Nugroho et al., 2023), yang mengemukakan bahwa durasi waktu yang lebih banyak dihabiskan untuk belajar memiliki dampak positif terhadap prestasi mahasiswa.

Selain itu, kesehatan mental, yang memiliki skor importansi sebesar 0.198, terbukti berperan penting dalam prediksi prestasi akademik. Hasil ini mendukung penelitian (Elik et al., 2024), yang menunjukkan bahwa kondisi psikologis yang baik, seperti rendahnya tingkat stres dan kecemasan, dapat meningkatkan daya fokus dan kinerja akademik mahasiswa. Jam tidur (0.154) juga terbukti memiliki kontribusi signifikan terhadap prestasi akademik, sejalan dengan temuan (Sihombing & Siagian, 2025), yang menegaskan bahwa kualitas tidur yang baik berpengaruh langsung terhadap kemampuan kognitif mahasiswa. Persentase kehadiran kelas, dengan skor importansi 0.121, menunjukkan bahwa mahasiswa yang lebih aktif dalam menghadiri kelas cenderung memiliki prestasi yang lebih baik, sesuai dengan temuan (Radja et al., 2024) yang mengaitkan keterlibatan akademik secara langsung dengan keberhasilan belajar. Sebaliknya, aktivitas digital rata-rata (0.098) ditemukan memiliki dampak negatif terhadap prestasi akademik mahasiswa. Hal ini mendukung penelitian (Ciptandini & Prathivi, 2025), yang menunjukkan bahwa penggunaan media digital yang berlebihan dapat mengalihkan perhatian mahasiswa dari kegiatan belajar, sehingga menurunkan konsentrasi dan prestasi akademik mereka.

Secara keseluruhan, hasil penelitian ini menegaskan bahwa faktor perilaku seperti jam belajar dan kehadiran kelas memiliki pengaruh dominan terhadap prestasi akademik mahasiswa. Namun, faktor psikologis dan gaya hidup sehat, seperti kesehatan mental dan kualitas tidur, juga menunjukkan peran yang sangat signifikan. Sebaliknya, pengaruh negatif dari aktivitas digital harus mendapatkan perhatian lebih dalam merancang kebijakan akademik yang lebih komprehensif dan holistik. Temuan ini menggarisbawahi pentingnya pendekatan multidimensi dalam pembinaan akademik yang tidak hanya memperhatikan faktor akademik tradisional, tetapi juga kondisi psikologis dan gaya hidup mahasiswa.

4. KESIMPULAN

Penelitian ini berhasil membangun model klasifikasi prestasi akademik mahasiswa dengan menggunakan algoritma *Random Forest* yang dioptimasi dengan *Particle Swarm Optimization* (PSO). Hasil penelitian menunjukkan bahwa optimasi PSO secara signifikan meningkatkan akurasi model, dari 86,7% menjadi 89,4%, serta memperbaiki metrik evaluasi lainnya seperti *presisi*, *recall*, dan *F1-score*. Analisis *feature importance* mengidentifikasi lima fitur yang paling berpengaruh terhadap prediksi prestasi akademik mahasiswa, yaitu jam belajar per hari, kesehatan mental, jam tidur, kehadiran kelas, dan aktivitas digital. Fitur jam belajar per hari dan kesehatan mental terbukti memiliki kontribusi terbesar, yang mengindikasikan bahwa faktor perilaku dan psikologis memainkan peran utama dalam prestasi akademik. Di sisi lain, aktivitas digital ditemukan memiliki dampak negatif, yang menunjukkan pentingnya pengelolaan waktu digital bagi mahasiswa untuk mencegah distraksi yang dapat mengurangi efektivitas belajar. Meskipun demikian, penelitian ini memiliki beberapa keterbatasan, antara lain jumlah data yang digunakan relatif terbatas serta belum mempertimbangkan faktor-faktor eksternal seperti motivasi intrinsik dan latar belakang ekonomi keluarga secara mendalam. Selain itu, model ini menggunakan data cross-sectional, sehingga belum mampu menangkap dinamika perubahan prestasi mahasiswa dari waktu ke waktu. Untuk penelitian selanjutnya, disarankan untuk menggunakan data longitudinal yang lebih besar serta melibatkan variabel yang lebih komprehensif, termasuk aspek psikososial dan ekonomi. Penggunaan algoritma lain seperti *deep learning* atau pengembangan model prediktif yang lebih kompleks juga dapat dieksplorasi guna meningkatkan akurasi dan cakupan analisis. Selain itu, pendekatan interpretasi model seperti SHAP (*SHapley Additive exPlanations*) juga dapat dikembangkan lebih lanjut agar hasil visualisasi dan interpretasi model dapat lebih representatif dan komunikatif bagi pemangku kepentingan (Puspanagara, 2025). Secara keseluruhan, penelitian ini memberikan kontribusi dalam pengembangan model prediktif berbasis machine learning di bidang pendidikan, dengan menunjukkan bahwa pendekatan seperti *Random Forest* dapat menjadi alat yang efektif dalam memprediksi prestasi akademik mahasiswa serta memberikan wawasan untuk intervensi pendidikan yang lebih tepat sasaran.

REFERENCES

- Basri, H., Azis, M. S., Malau, Y., & Fridayanthie, E. W. (2022). Penerapan Particle Swarm Optimization Pada Algoritma Naïve Bayes Untuk Klasifikasi Hasil Belajar. *Information System For Educators And Professionals*, 6(2), 97–106. file:///D:/Kuliah S2

- UTI/Semester 2 2024-2025/Data Sains/Seminar Nasional/Refrensi/S4/1752-61-6398-1-10-20220802.pdf
- Ciptandini, M. I., & Prathivi, R. (2025). Klasterisasi Tingkat Kecanduan Penggunaan Tiktok Terhadap Minat Belajar Menggunakan Algoritma K-Means Clustering. *KESATRIA : Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)*, 6(1), 77–84. file:///D:/Kuliah S2 UTI/Semester 2 2024-2025/Data Sains/Seminar Nasional/Refrensi/S4/548-1094-1-SM.pdf
- Elik, H. K., Ama, R. G. T., Evianawati, E., Bete, R. N. S., & S. Bete, R. N. (2024). Dukungan Sosial Orang Tua, Hardiness Dan Stres Akademik Pada Mahasiswa Tingkat Akhir. *DE JOURNAL (Dharmas Education Journal)*, 5(1), 304–313. <https://doi.org/10.56667/dejournal.v5i1.1311>
- Hariyanti, I., Al-husaini, M., Raharja, A. R., Husaini, A., Hariyanti, I., & Raharja, A. R. (2024). Perbandingan Algoritma Decision Tree dan Naive Bayes dalam Klasifikasi Data Pengaruh Media Sosial dan Jam Tidur Terhadap Prestasi Akademik Siswa. *Technologia : Jurnal Ilmiah*, 15(2), 332–340. <https://doi.org/10.31602/tji.v15i2.14381>
- Hasibuan, I. I. S. (2024). Pengaruh Kurang Tidur Terhadap Short Term Memory Pada Mahasiswa Kedokteran Universitas Malikussaleh. 5(4), 9–10.
- Jayanta Nath. (2025). *Dataset student_habits_performance*. Kaggle. <https://www.kaggle.com/datasets/jayaantanaath/student-habits-vs-academic-performance>
- Kuswanto, J., & Hakim, L. (2025). Penerapan Algoritma Random Forest untuk memprediksi Performa Akademik Mahasiswa. *DECODE : Jurnal Pendidikan Teknologi Informasi*, 5(1), 262–270. file:///D:/Kuliah S2 UTI/Semester 2 2024-2025/Data Sains/Seminar Nasional/Refrensi/S3/262-270-artikel-1103-publish.pdf
- Ma'ruf, S., Nurkadri, Sitopu, G. S., & Habeahan, G. F. (2024). Hubungan antara Olahraga dan Kesehatan Mental Saufi. *Jurnal Cerdas SIFA Pendidikan*, 13(1), 1–15. file:///D:/Kuliah S2 UTI/Semester 2 2024-2025/Data Sains/Seminar Nasional/Refrensi/S5/e_yuliawan,+33728-Article+Text-100269-1-6-20240528+(1).pdf
- Mawaddah, S., Pranoto, W. J., & Faldi. (2023a). Optimasi Algoritma C4 . 5 Menggunakan Metode Adaboost Classification Pada Klasifikasi Nilai Mahasiswa Studi Kasus : Universitas Muhammadiyah Kalimantan Timur Optimizing the C4 . 5 Algorithm Using Adaboost Classification Method for Student Grade Classifica. *Jurnal Sains Komputer Dan Teknologi Informasi*, 6(1), 83–89. file:///D:/Kuliah S2 UTI/Semester 2 2024-2025/Data Sains/Seminar Nasional/Refrensi/S458-Article Text-23790-1-10-20231218.pdf
- Mawaddah, S., Pranoto, W. J., & Faldi, F. (2023b). Optimasi Algoritma C4.5 Menggunakan Metode Adaboost Classification Pada Klasifikasi Nilai Mahasiswa Studi Kasus: Universitas Muhammadiyah Kalimantan Timur. *Jurnal Sains Komputer Dan Teknologi Informasi*, 6(1), 83–89. <https://doi.org/10.33084/jsakti.v6i1.5458>
- Nugroho, B. I., Santoso, N. A., Murtopo, A. A., Korespondensi, P., & Pendahuluan, I. (2023). Prediksi Kemampuan Akademik Mahasiswa dengan Metode Support Vector Machine. *Remik : Riset Dan E-Jurnal Manajemen Informatika Komputer*, 7(1), 177–188. <https://doi.org/10.33395/remik.v7i1.12010>
- Nurhidayat, A. I., & Fatrianto, D. (2021). *Prediksi Kinerja Akademik Mahasiswa Menggunakan Machine Learning dengan Sequential Minimal Optimization untuk Pengelola Program Studi*. 05, 84–91.
- Pratama, R. S., & Zaimah, A. F. N. (2024). Pengaruh Aktivitas Fisik Terhadap Kesehatan Mental Dan Kinerja Akademik Mahasiswa Pendidikan Bahasa Inggris. *Jurnal Lingkar Pembelajaran Inovatif*, 5(9), 94–104. file:///D:/Kuliah S2 UTI/Semester 2 2024-2025/Data Sains/Seminar Nasional/Refrensi/S6/94-104.pdf
- Puspanagara, A. L. (2025). Penerapan Explainable AI untuk Prediksi Performa Akademik Mahasiswa Menggunakan Random Forest dan SHAP. *Infoman's: Jurnal Ilmu-Ilmu Informatika Dan Manajemen*, 19(1), 1–7. <https://doi.org/10.13140/RG.2.2.27853.14565>
- Radja, Y. L., Limbu, R., Bunga, E. Z. H., & Weraman, P. (2024). Analisis Hubungan Smartphone Addiction , Kecemasan , dan Stres Terhadap Kejadian Insomnia pada Mahasiswa Program Studi Kesehatan Masyarakat Universitas Nusa Cendana Kupang. *SEHATMAS (Jurnal Ilmiah Kesehatan Masyarakat)*, 3(4), 798–813. <https://doi.org/10.55123/sehatmas.v3i4.4169>
- Rahman, D. (2023). Analisis gaya hidup sehat mahasiswa olahraga. *Jurnal Patriot*, 5(3), 239–246. <https://doi.org/10.24036/patriot.v5i3.1018>
- Sihombing, A. S., & Siagian, N. (2025). Hubungan Kualitas Tidur Terhadap Indeks Prestasi Kumulatif Mahasiswa Keperawatan Di Universitas Advent Indonesia. *Syntax Literate: Jurnal Ilmiah Indonesia*, 10(4), 4359–4374. file:///D:/Kuliah S2 UTI/Semester 2 2024-2025/Data Sains/Seminar Nasional/Refrensi/S4/58757-Article Text-72234-1-10-20250502.pdf
- siska Fitriani. (2025). *semmas*. <https://colab.research.google.com/drive/1yYxKylaGCGbhhLI0RyFNX-CDiSO4rzw7#scrollTo=dEjWZpbcxk4U>
- Wijyaningrum, V. N., Putri, I. K., Kirana, A. P., Mubarak, M. R., Harahap, D. M., & Hamesha, B. R. (2023). Analisis performa seleksi atribut untuk menentukan potensi mahasiswa putus studi. *JIP (Jurnal Informatika Polinema)*, 9(2), 237–244. <https://doi.org/10.33795/jip.v9i2.1300>
- Zahir, A., Nur, H., Hidayat, W., & Parubang, D. (2021). Evaluasi Hasil Belajar Elektronika Digital melalui Tes Formatif , Sumatif , dan Remedial. *Jurnal Literasi Digital*, 1(2), 122–129. file:///D:/Kuliah S2 UTI/Semester 2 2024-2025/Data Sains/Seminar Nasional/Refrensi/13-Article Text-154-1-10-20210906.pdf