

Prediksi Performa Akademik Mahasiswa Berdasarkan Gaya Hidup dan Dukungan Sosial Menggunakan Random Forest

Royan Ristu Prayoga¹⁾, Anna Baita, S.Kom., M.Kom.²⁾

¹⁾ Informatika, Universitas Amikom Yogyakarta

²⁾ Informatika, Universitas Amikom Yogyakarta

email : royanristuprayoga@students.amikom.ac.id¹⁾, email : anna@mikom.ac.id²⁾

Abstraksi

Performa akademik mahasiswa dipengaruhi oleh berbagai faktor, termasuk gaya hidup dan dukungan sosial yang diterima selama proses studi. Penelitian ini bertujuan untuk memprediksi performa akademik mahasiswa dengan memanfaatkan metode data mining, khususnya algoritma Random Forest, yang terbukti efektif dalam menghasilkan prediksi yang akurat. Data yang digunakan merupakan hasil survei terkait kebiasaan hidup mahasiswa dan tingkat dukungan sosial yang diperoleh. Proses penelitian meliputi tahap preprocessing data, pemberian label pada atribut non-numerik, dan validasi model menggunakan teknik cross-validation. Model dievaluasi menggunakan metrik Confusion Matrix, termasuk accuracy, precision, recall, dan F1-score. Hasil eksperimen menunjukkan bahwa algoritma Random Forest mampu mengklasifikasikan performa akademik mahasiswa dengan akurasi yang tinggi, serta memberikan informasi yang berguna bagi pihak kampus dalam pengambilan keputusan strategis untuk meningkatkan kualitas pembelajaran dan dukungan terhadap mahasiswa.

Kata Kunci :

performa akademik, gaya hidup, dukungan sosial, data mining, Random Forest

Abstract

Student academic performance is influenced by various factors, including lifestyle and the social support received during their studies. This research aims to predict student academic performance using data mining methods, specifically the Random Forest algorithm, which has proven effective in generating accurate predictions. The data used is derived from surveys regarding student lifestyle habits and the level of social support obtained. The research process includes data preprocessing, labeling non-numeric attributes, and model validation using cross-validation techniques. The model is evaluated using Confusion Matrix metrics, including accuracy, precision, recall, and F1-score. Experimental results indicate that the Random Forest algorithm can classify student academic performance with high accuracy, providing valuable information for the campus in making strategic decisions to improve the quality of learning and support for students.

Keywords :

Academic performance, Lifestyle, Social support, Data mining, Random Forest

1. Pendahuluan

Tingkat kelulusan atau graduation rate merupakan salah satu indikator utama dalam mengukur kinerja sebuah lembaga akademik [1]. Di berbagai negara, termasuk Indonesia, angka putus kuliah masih menjadi tantangan serius yang berdampak pada masa depan generasi muda. Studi menunjukkan bahwa lebih dari 15% mahasiswa di negara-negara maju seperti Amerika Serikat, Jerman, dan Inggris mengalami dropout selama masa studi mereka [1]. Di Indonesia sendiri, tingkat mahasiswa yang melanjutkan ke jenjang perguruan tinggi masih rendah, yaitu hanya sekitar 6% dari total lulusan SMA, dengan tingkat putus kuliah yang cukup tinggi di berbagai provinsi, seperti Bengkulu dan Jawa Timur [3].

Penelitian sebelumnya telah menunjukkan bahwa rendahnya performa akademik merupakan salah satu indikator utama dari tingginya angka putus kuliah [1][2]. Oleh karena itu, penting bagi institusi pendidikan untuk memahami faktor-faktor yang memengaruhi keberhasilan akademik mahasiswa guna merumuskan strategi pencegahan yang efektif. Faktor-faktor tersebut mencakup kondisi sosial ekonomi, gaya hidup, dukungan sosial, hingga variabel personal seperti stres, kebiasaan belajar, dan kesehatan [1][3].

Seiring berkembangnya teknologi, pendekatan berbasis Educational Data Mining (EDM) dan algoritma machine learning seperti Decision Tree, Naïve Bayes, Random Forest, dan K-Nearest Neighbor telah banyak digunakan untuk memprediksi performa akademik mahasiswa [2][3]. Di antara

algoritma tersebut, Random Forest memiliki keunggulan dalam hal akurasi dan stabilitas prediksi pada data pendidikan yang kompleks [3]. Berbagai penelitian menunjukkan bahwa Random Forest mampu menghasilkan akurasi hingga 91% dalam memprediksi keberhasilan akademik [3], menjadikannya salah satu algoritma yang efektif untuk digunakan dalam sistem pendukung keputusan pendidikan.

Namun demikian, sebagian besar penelitian sebelumnya lebih banyak menitikberatkan pada variabel akademik seperti IPK dan absensi, tanpa memasukkan faktor gaya hidup dan dukungan sosial secara menyeluruh sebagai variabel penting dalam model prediksi [1][3]. Padahal, aspek-aspek tersebut memiliki peran signifikan dalam menentukan keberhasilan studi mahasiswa.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengembangkan model prediksi performa akademik mahasiswa dengan mempertimbangkan faktor gaya hidup dan dukungan sosial menggunakan algoritma Random Forest. Dengan pendekatan ini, diharapkan institusi pendidikan dapat memperoleh wawasan yang lebih luas dalam memahami karakteristik mahasiswa dan menyusun kebijakan yang lebih tepat sasaran untuk meningkatkan kualitas pembelajaran dan menekan angka putus kuliah.

2. Metode Penelitian

2.1 Jenis Penelitian

Ini adalah penelitian kuantitatif prediktif yang menggunakan data mining dan machine learning untuk memprediksi performa akademik mahasiswa. Tujuan utamanya adalah membangun model yang bisa memperkirakan performa ini berdasarkan data gaya hidup dan dukungan sosial, dengan memanfaatkan algoritma Random Forest Regressor.

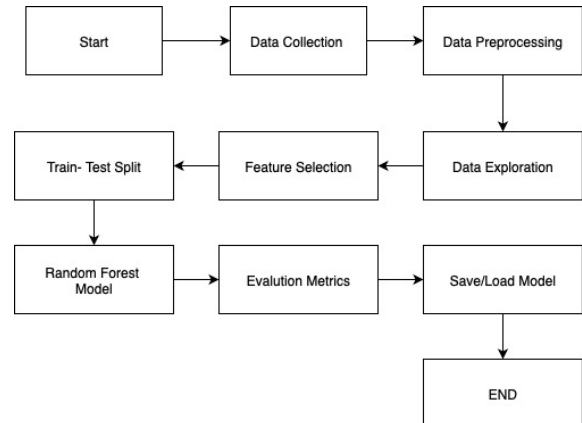
Penelitian ini menggunakan pendekatan kuantitatif karena semua tahapannya melibatkan pengumpulan, pemrosesan, dan analisis data numerik (kuantitatif) untuk menjawab pertanyaan penelitian secara statistik dan sistematis. Sifat prediktifnya cocok karena fokus utama penelitian ini adalah mengembangkan model untuk memprediksi performa akademik dari variabel-variabel yang sudah dikumpulkan[2].

2.2 Alur Penelitian

Penelitian ini diawali dengan pengumpulan data sekunder dari dataset `enhanced_student_habits_performance_dataset.csv`, yang memuat informasi gaya hidup, kebiasaan belajar, dan faktor sosial mahasiswa. Data kemudian diproses melalui penghapusan atribut tidak relevan, penanganan nilai kosong dan duplikat, serta transformasi data kategorikal menggunakan Label

Encoding. Fitur numerik distandarisasi untuk meningkatkan kinerja model.

Setelah dilakukan pemisahan fitur (X) dan target (y), data dibagi menjadi data latih dan data uji dengan rasio 80:20. Model yang digunakan adalah Random Forest Regressor dengan parameter utama: 200 pohon keputusan (`n_estimators`) dan kedalaman maksimum 10. Model dilatih pada data latih dan dievaluasi menggunakan metrik MAE, MSE, RMSE, dan R^2 [2].



Gambar 1. Tahap Penelitian

2.3 Pengambilan Dataset

Data yang dimanfaatkan dalam penelitian ini berasal dari Kaggle, sebuah platform sumber data publik, di bawah judul "Enhanced Student Habits Performance Dataset." Dataset ini dirancang khusus untuk memfasilitasi analisis dan prediksi performa akademik mahasiswa, dengan mempertimbangkan beragam faktor seperti kebiasaan belajar, gaya hidup, dan dukungan sosial. Dataset tersebut dapat diakses secara bebas melalui tautan: <https://www.kaggle.com/datasets/aljarah/xAPI-Edu-Data>. Fitur-fitur tersebut pada dataset ini meliputi IPK sebelumnya (previous GPA), persentase kehadiran, rata-rata jam belajar per hari, tingkat stres, waktu penggunaan perangkat elektronik (screen time), dan skor kemampuan manajemen waktu (time management score). Variabel-variabel ini berperan sebagai prediktor (variabel independen) untuk memodelkan nilai ujian (exam_score), yang merupakan representasi dari performa akademik mahasiswa sebagai variabel target. Setiap entri atau baris dalam dataset ini merepresentasikan satu individu mahasiswa, lengkap dengan kombinasi nilai dari seluruh fitur yang disebutkan. Dataset telah melalui serangkaian praproses, termasuk penghapusan data duplikat dan penanganan nilai yang hilang. Selain itu, variabel kategorikal telah diubah menggunakan label encoding, dan data numerik telah dinormalisasi dengan Standard Scaler untuk mengoptimalkan kinerja model. Detail mengenai setiap fitur dalam dataset akan dijelaskan lebih lanjut dalam Tabel 1.

Tabel 1. Dataset

Fitur	Deskripsi	Tipe Data
previous_gpa	IPK mahasiswa sebelumnya	Numerik
attendance_percentage	Persentase kehadiran mahasiswa	Numerik
study_hours_per_day	Rata-rata waktu belajar per hari (jam)	Numerik
stress_level	Tingkat stres mahasiswa	Numerik
screen_time	Waktu penggunaan layar per hari (jam)	Numerik
time_management_score	Skor kemampuan manajemen waktu	Numerik
exam_score	Skor ujian akhir mahasiswa (variabel target)	Numerik

2.4 Preprocessing

Preprocessing merupakan tahap awal yang penting dalam proses pengolahan data, dengan tujuan untuk membersihkan dan mempersiapkan data agar siap digunakan dalam tahap modeling selanjutnya [13]. Pada penelitian ini, preprocessing dilakukan dilakukan pemeriksaan terhadap data hilang (missing values) dan data duplikat. Baris dengan nilai kosong pada variabel target (exam_score) dihapus, untuk menghindari ketidaksesuaian saat proses pelatihan model. Penanganan missing value dilakukan karena informasi yang hilang dapat menyebabkan bias dalam analisis dan penurunan akurasi prediksi [14].

Kemudian, proses encoding dilakukan terhadap atribut bertipe kategorikal menggunakan metode Label Encoding, agar dapat diubah menjadi bentuk numerik yang dapat dibaca oleh algoritma pembelajaran mesin. Hal ini diperlukan karena model seperti Random Forest hanya dapat bekerja dengan data numerik. Langkah terakhir adalah normalisasi data menggunakan metode Standard Scaler, yang mengubah fitur numerik agar memiliki distribusi dengan mean = 0 dan standar deviasi = 1. Normalisasi ini dilakukan untuk menghindari dominasi fitur yang memiliki skala nilai lebih besar terhadap proses pembelajaran model. Dengan normalisasi, skala setiap fitur menjadi seragam, sehingga membantu mencegah bias dan meningkatkan stabilitas model [15].

2.5 Data Exploration

Eksplorasi data bertujuan untuk memahami pola dan hubungan antar variabel dalam dataset. Distribusi fitur-fitur seperti previous_gpa, attendance_percentage, study_hours_per_day, dan lainnya divisualisasikan menggunakan histogram, yang menunjukkan variasi data yang beragam. Korelasi antar variabel dihitung dan divisualisasikan dengan heatmap, di mana previous_gpa dan

attendance_percentage menunjukkan korelasi positif terhadap exam_score, sedangkan stress_level dan screen_time cenderung negatif. Model Random Forest digunakan untuk mengevaluasi tingkat kepentingan fitur, dan hasilnya menunjukkan bahwa previous_gpa dan study_hours_per_day merupakan fitur paling berpengaruh. Scatter plot antara nilai actual dan prediksi memperlihatkan kecenderungan prediksi model mengikuti data aktual, menandakan performa model yang baik.

2.6 Feature Selection

Feature selection atau pemilihan fitur merupakan langkah penting untuk meningkatkan performa model dengan hanya menggunakan variabel yang paling relevan. Dalam penelitian ini, teknik feature importance dari algoritma Random Forest digunakan untuk mengidentifikasi fitur-fitur yang memiliki kontribusi paling besar dalam memprediksi exam_score. Hasil seleksi menunjukkan bahwa fitur previous_gpa, attendance_percentage, study_hours_per_day, stress_level, screen_time, dan time_management_score merupakan variabel yang paling signifikan dan dipilih sebagai input untuk model prediktif. Pemilihan fitur ini tidak hanya menyederhanakan model, tetapi juga membantu mengurangi kompleksitas komputasi dan risiko overfitting.

2.7 Train- Test Split

Tahap ini bertujuan untuk membagi data menjadi dua bagian utama, yaitu data latih (training data) dan data uji (testing data). Pada penelitian ini, pembagian dilakukan dengan proporsi 80% data untuk pelatihan model dan 20% untuk pengujian, menggunakan fungsi train_test_split dari pustaka Scikit-Learn. Teknik ini penting untuk memastikan bahwa model belajar dari sebagian data dan dievaluasi pada data yang belum pernah dilihat sebelumnya. Dengan demikian, diperoleh gambaran yang objektif mengenai performa model saat diterapkan pada data nyata. Proses ini dilakukan setelah tahapan pembersihan dan transformasi data selesai dilakukan.

$$|D_{train}| = \alpha \cdot |D|, |D_{test}| = (1-\alpha) \cdot |D|$$

- D adalah himpunan data awal (dataset) yang berjumlah n data,
- D train adalah himpunan data latih,
- D test adalah himpunan data uji,
- α adalah rasio data latih terhadap keseluruhan data (dalam penelitian ini: 0.8 atau 80%),
- $(1-\alpha)$ adalah rasio data uji (dalam penelitian ini: 0.2 atau 20%).

2.8 Random Forest Model

Random Forest merupakan algoritma ensemble learning yang digunakan dalam penelitian ini untuk melakukan prediksi nilai performa akademik mahasiswa berdasarkan beberapa variabel gaya hidup dan dukungan sosial. Algoritma ini bekerja dengan

membangun sejumlah pohon keputusan (decision trees) pada data pelatihan, dan menghasilkan output prediksi berdasarkan rata-rata (pada regresi) dari hasil masing-masing pohon tersebut [16].

Model Random Forest yang digunakan dalam penelitian ini dibangun dengan parameter $n_estimators=200$, $max_depth=10$, dan $random_state=42$ untuk memastikan reproduisibilitas hasil. Proses pelatihan model dilakukan setelah melalui tahap pra-pemrosesan dan pembagian data ke dalam subset data latih dan data uji. Hasil dari model ini menunjukkan tingkat akurasi prediksi yang baik, yang ditunjukkan oleh metrik evaluasi seperti Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), dan R-squared (R^2).

Secara matematis, Random Forest untuk regresi dapat direpresentasikan sebagai:

$$y = \frac{1}{T} \sum_{t=1}^T h_t(x)$$

- y adalah nilai prediksi akhir
- T adalah jumlah pohon dalam hutan (jumlah estimator)
- $h_t(x)$ adalah prediksi dari pohon ke- t
- x adalah vektor fitur input.

2.9 Evaluation Metrics

Evaluasi performa model dalam penelitian ini dilakukan dengan menggunakan beberapa metrik regresi, yaitu Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), dan R-squared (R^2). MAE mengukur rata-rata selisih absolut antara nilai prediksi dan nilai aktual, sedangkan MSE dan RMSE memberikan penalti lebih besar terhadap kesalahan yang lebih jauh. R-squared (R^2) digunakan untuk menilai seberapa besar variasi nilai target yang dapat dijelaskan oleh model. Hasil evaluasi menunjukkan bahwa model Random Forest Regressor mampu memberikan prediksi nilai ujian mahasiswa dengan tingkat akurasi yang baik. Selain itu, visualisasi plot antara nilai aktual dan nilai prediksi yang telah dibuat menggambarkan sebaran titik data yang mendekati garis diagonal, yang mengindikasikan bahwa hasil prediksi cukup sesuai dengan data sebenarnya.

3. Hasil dan Pembahasan

Pada bagian ini akan dijelaskan hasil-hasil yang diperoleh dari proses Prediksi Performa Akademik Mahasiswa Berdasarkan Gaya Hidup dan Dukungan Sosial. Hasil penelitian disusun secara sistematis mulai dari evaluasi model, visualisasi data, pembahasan performa algoritma, hingga kelebihan dan kekurangan sistem.

3.1 Hasil Evaluasi Model

Pasca selesainya tahapan pra-pemrosesan dan seleksi fitur, kami melanjutkan pada proses pelatihan model.

Untuk memastikan robustnya hasil, setiap model dilatih menggunakan teknik validasi silang (cross-validation), sebelum kemudian diuji secara menyeluruh menggunakan data uji yang telah dialokasikan. Kami melakukan evaluasi performa model secara cermat, memanfaatkan serangkaian metrik klasifikasi standar, meliputi akurasi, presisi, recall, dan F1-score, guna menilai efektivitas prediksi secara komprehensif. Hasil evaluasi disajikan pada Tabel 1.

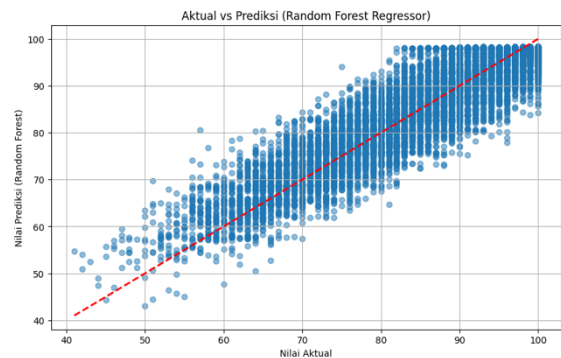
Tabel 2. Hasil Evaluasi Model

Model	MAE	MSE	RMSE	R^2
Random Forest	3.2344	17.4539	4.1778	0.8711

Hasil evaluasi ini menunjukkan bahwa model mampu memprediksi nilai ujian akhir mahasiswa dengan tingkat akurasi yang cukup tinggi, ditunjukkan oleh nilai R^2 sebesar 0.8711. Artinya, sekitar 87.1% variasi dari nilai ujian mahasiswa dapat dijelaskan oleh model berdasarkan fitur gaya hidup dan dukungan sosial yang digunakan.

3.2 Visualisasi Hasil Prediksi

Untuk mengevaluasi kinerja model Random Forest Regressor, dilakukan visualisasi antara nilai aktual ($exam_score$) dan nilai prediksi yang dihasilkan oleh model.

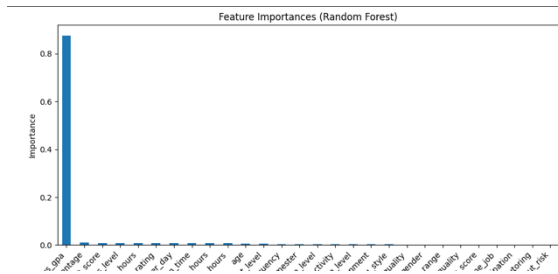


Gambar 2. Aktual vs Prediksi

Gambar 3 menunjukkan hubungan antara nilai aktual (sumbu X) dan prediksi model (sumbu Y). Titik-titik biru mewakili pasangan nilai aktual dan hasil prediksi dari data uji. Garis putus-putus berwarna merah merupakan garis identitas ($y = x$), yang menunjukkan prediksi sempurna. Sebagian besar titik-titik dalam scatter plot cenderung berada di sekitar garis merah, yang mengindikasikan bahwa model mampu memprediksi nilai ujian mahasiswa dengan tingkat akurasi yang baik. Meskipun terdapat beberapa outlier, distribusi prediksi menunjukkan pola linier yang jelas dan sejalan dengan nilai aktualnya. Hal ini juga didukung oleh nilai R^2 sebesar 0.825, yang berarti 82.5% variasi nilai ujian dapat dijelaskan oleh model. Visualisasi ini memperkuat kesimpulan bahwa model Random Forest yang dibangun cukup handal dalam menangkap hubungan antara gaya hidup dan dukungan sosial terhadap performa akademik mahasiswa.

3.3 Visualisasi Feature Importance

Salah satu keunggulan algoritma Random Forest adalah kemampuannya dalam mengevaluasi seberapa besar kontribusi masing-masing fitur terhadap hasil prediksi. Visualisasi feature importance membantu dalam mengidentifikasi variabel-variabel yang paling relevan.

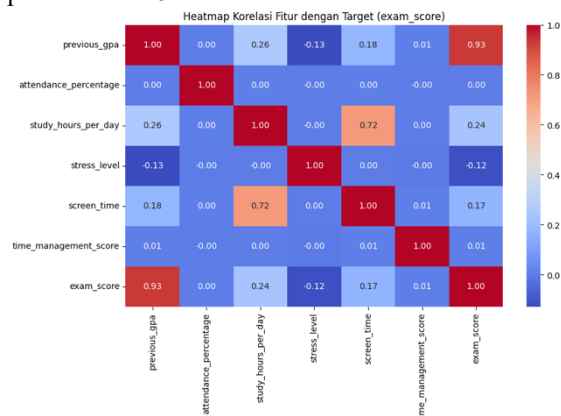


Gambar 3. Feature Importances

Gambar 4 menunjukkan bahwa fitur `previous_gpa` memberikan pengaruh yang sangat dominan terhadap hasil prediksi, dengan bobot penting di atas 0.85. Fitur ini secara jelas menjadi indikator utama dalam memprediksi `exam_score`. Sementara fitur lainnya seperti `attendance_percentage`, `study_hours_per_day`, `time_management_score`, dan `stress_level` memiliki nilai penting yang jauh lebih kecil, meskipun tetap berkontribusi dalam pembentukan model secara keseluruhan. Temuan ini sejalan dengan konsep akademik bahwa riwayat akademik (IPK sebelumnya) menjadi salah satu indikator kuat dalam memprediksi performa masa depan. Namun demikian, fitur gaya hidup seperti waktu belajar, stres, dan manajemen waktu tetap penting dalam konteks personalisasi dan intervensi kebijakan pendidikan.

3.4 Korelasi Antar Variabel

Untuk memahami hubungan antara fitur-fitur gaya hidup dan target prediksi (`exam_score`), dilakukan analisis korelasi menggunakan metode Pearson. Hasilnya divisualisasikan dalam bentuk heatmap pada Gambar 5.



Gambar 5. Heatmap Korelasi Fitur dengan Target
Berdasarkan Gambar 5, terdapat beberapa temuan penting:

- `previous_gpa` memiliki korelasi sangat kuat terhadap `exam_score` ($r = 0.93$), yang

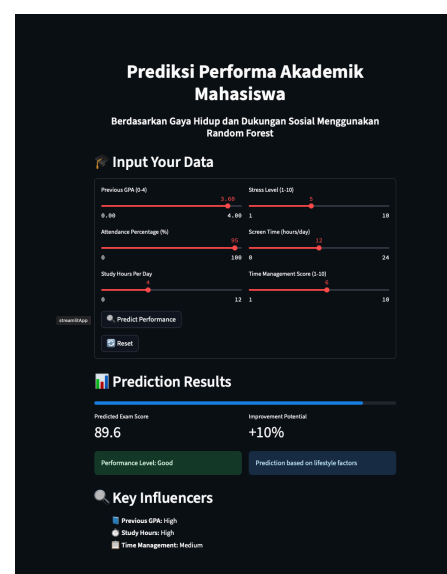
memperkuat hasil analisis feature importance sebelumnya.

- `study_hours_per_day` menunjukkan korelasi positif moderat terhadap `exam_score` ($r = 0.24$), yang menandakan semakin banyak waktu belajar, semakin tinggi kemungkinan nilai ujian meningkat.
- Variabel lain seperti `stress_level` dan `screen_time` memiliki korelasi negatif lemah terhadap `exam_score`, mengindikasikan potensi pengaruh negatif terhadap performa akademik, namun tidak terlalu dominan.
- Korelasi antar fitur juga terdeteksi, seperti antara `screen_time` dan `study_hours_per_day` ($r = 0.72$), yang menunjukkan bahwa waktu layar cenderung berhubungan dengan kebiasaan belajar.

Analisis korelasi ini memberikan wawasan awal terhadap hubungan linear antara variabel, namun tidak menggambarkan hubungan non-linear atau interaksi kompleks, yang justru bisa ditangkap oleh model Random Forest.

3.5 Deploy Model

Setelah proses pemodelan dan evaluasi selesai dilakukan, tahap selanjutnya adalah melakukan deployment model agar dapat digunakan secara langsung oleh pengguna. Tujuan dari deployment ini adalah memungkinkan pengguna—seperti dosen, pihak akademik, atau mahasiswa sendiri—untuk memprediksi nilai ujian (`exam_score`) berdasarkan input variabel gaya hidup dan dukungan sosial yang dimiliki mahasiswa. Model Random Forest yang telah dilatih sebelumnya disimpan dalam format `.pkl` menggunakan library Python seperti `joblib` atau `pickle`. Model ini kemudian dapat diintegrasikan ke dalam aplikasi berbasis web atau antarmuka pengguna sederhana seperti Streamlit, Flask, atau Jupyter Notebook interaktif. Berikut adalah tampilan antarmuka pengguna ketika model digunakan untuk memprediksi nilai ujian berdasarkan input pengguna



Gambar 6. Tampilan Antarmuka

Antarmuka ini dirancang agar intuitif dan mudah digunakan oleh semua kalangan. Selain memberikan prediksi numerik (seperti nilai prediksi 89.6), aplikasi juga menyediakan insight tambahan seperti potensi perbaikan (+10%) dan interpretasi performa (misalnya “Good”), yang akan sangat membantu dalam proses pengambilan keputusan terkait strategi belajar mahasiswa.

4. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan mengenai Prediksi Performa Akademik Mahasiswa Berdasarkan Gaya Hidup dan Dukungan Sosial Menggunakan Random Forest, maka dapat diambil beberapa kesimpulan sebagai berikut:

- Antarmuka ini dirancang agar intuitif dan mudah digunakan oleh semua kalangan. Selain memberikan prediksi numerik (seperti nilai prediksi 89.6), aplikasi juga menyediakan insight tambahan seperti potensi perbaikan (+10%) dan interpretasi performa (misalnya “Good”), yang akan sangat membantu dalam proses pengambilan keputusan terkait strategi belajar mahasiswa.
- Fitur previous_gpa merupakan faktor yang paling dominan dalam memengaruhi performa akademik mahasiswa, diikuti oleh attendance_percentage, study_hours_per_day, dan time_management_score. Hal ini menunjukkan bahwa riwayat akademik sebelumnya serta kebiasaan belajar harian memainkan peran penting dalam pencapaian nilai akhir mahasiswa.
- Visualisasi hasil prediksi menunjukkan distribusi titik data yang cenderung mendekati garis identitas, mengindikasikan bahwa model memberikan prediksi yang cukup akurat dan stabil. Selain itu, hasil korelasi juga mengonfirmasi adanya hubungan kuat antara beberapa variabel gaya hidup dengan performa akademik mahasiswa.
- Model yang telah dibangun telah berhasil dideploy ke dalam antarmuka pengguna yang sederhana dan intuitif. Hal ini memungkinkan stakeholder seperti dosen dan mahasiswa untuk memanfaatkan model prediktif secara langsung dalam pengambilan keputusan akademik maupun strategi belajar yang lebih baik.
- Penelitian ini menunjukkan bahwa integrasi antara data gaya hidup dan dukungan sosial dengan pendekatan pembelajaran mesin seperti Random Forest dapat memberikan kontribusi signifikan dalam memprediksi dan meningkatkan performa akademik mahasiswa. Namun demikian, terdapat ruang untuk pengembangan lebih lanjut dengan mempertimbangkan variabel tambahan dan eksplorasi model yang lebih kompleks.

Daftar Pustaka

- [1] S. Rajendran, S. Chamundeswari, and A. A. Sinha, “Predicting the academic performance of middle- and high-school students using machine learning

algorithms,” *Social Sciences & Humanities Open*, vol. 6, 100357, 2022.

[2] S. Linawati, S. Nurdiani, K. Handayani, and Latifah, “Prediksi prestasi akademik mahasiswa menggunakan algoritma Random Forest dan C4.5,” *Jurnal Khatulistiwa Informatika*, vol. VIII, no. 1, pp. –, Jun. 2020.

[3] K. Afandi, M. H. Arief, N. F. Laily, and D. M. Nugroho, “Analisis performa akademik mahasiswa menggunakan Social Network Analysis,” *Bisnis Digital*, Universitas dr. Soebandi, Jember, Indonesia.

[4] S.-S. M. Ajibade, J. Dayupay, D.-L. Ngo-Hoang, O. J. Oyeboode, J. M. Sasan, N. G. Gido, A. C. Tabuena, and A. Adediran, “Utilization of ensemble techniques for prediction of the academic performance of students,” *Journal of Optoelectronics Laser*, vol. 41, no. 6, pp. –, 2022.

[5] Z. A. Dwiyantri and C. Prianto, “Prediksi cuaca Kota Jakarta menggunakan metode Random Forest: Studi optimalitas,” *Jurnal Tekno Insentif*, vol. 17, no. 2, pp. 127–137, Oct. 2023, doi: 10.36787/jti.v17i2.1136.

[6] B. Prasjo and E. Haryatmi, “Analisa prediksi kelayakan pemberian kredit pinjaman dengan metode Random Forest,” *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 7, no. 2, pp. 79–89, 2021. [Online]. Available: <http://teknosi.fti.unand.ac.id/>

[7] A. M. A. Rahim, I. Y. R. Pratiwi, and M. A. Fikri, “Klasifikasi penyakit jantung menggunakan metode Synthetic Minority Over-Sampling Technique dan Random Forest Classifier,” *Indonesian Journal of Computer Science*, pp. xx–xx. [Online]. Available: <https://ijcs.stmikindonesia.ac.id/>